


```

----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-656022 (sc0182)
----- Summary Statistics
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180536 bases at least Q40
Consensus quality: 180650 bases at least Q30
Consensus quality: 180657 bases at least Q20
Insert size: 194815; 11.0% error; agarose-fp
Insert size: 180657; sum-of-ctrls
Quality coverage: 8.4x in Q20 bases; agarose-fp
Quality coverage: 9.0x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 180657: contig of 180657 bp in length.
Location/Qualifiers
1.180657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-656022"
/clone.lib="RPC1 human BAC library 11"
1.180657
/notice="assembly name: Contig19"
-----
BASE COUNT 52363 a 34237 c 35512 g 58545 t
ORIGIN
misc_feature
1.180657
/notice="assembly name: Contig19"
-----
Query Match 99.8%; Score 922.4; DB 2; Length 180657;
Best Local Similarity 99.9%; Pred. No. 1.2e-251;
Matches 923; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
-----
QY 1 atgaatcacagcgtgtgaactgaagtcattatctctggagccctcaccaaaagcctgaactc 60
DB 155813 ATGAATCACAGCGTGTGAAGTCAATGATTTCTGGGCTCACAAAAGCCTGAACTC 155754
QY 61 cgggggaattatctctctctctctctctctctctctctctctctctctctctctctctct 120
DB 155753 CAGGGGAATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 155694
QY 121 ctcatacatcatgtccaaatctatagcaaacctctgcatagcccatatgcttctctt 180
DB 155693 CTCATCATCATTTGCCAAATCTATTAACAACACCTTGCAATGCCCATGATGATTTTCTT 155634
QY 181 ctgacactgctgtgtgtgacatcatctgcacaaagaagcatatccgaagatgctggg 240
DB 155633 CTGACACTGGCTGTGTGTGACATCATCTGCACAAAGCATCATACCGAATGCTGGGG 155574
QY 241 accatgtcaacatcacgaanaatacatcttatatgcaagctgcatgtccagctctcttg 300
DB 155573 ACATGTCAACATCACGAANAATACATTTATATGACAGGCTGATGCTCCAGCTCTCTTG 155514
QY 301 ttcaatgtctctgtggaagtgtgagttctctctcacacacatgacctatgacctat 360
DB 155513 TTCAATGTCTCTGTGGAAGTGTGAGTTCTCTTCAACACATGSCCATATGACCGCAT 155454
QY 361 gtggcattgttccctctcatctaaagtaattatgaaccacatatgtgtgagcc 420

```

```

DB 155453 GTGGCCATTTGTTCCCTCTCATTAACACTACTATTATGAAACCCATATGTGTAGCC 155394
QY 421 tgcctgaagctgtcatatgctctatgctgaatcaccatctctgggtgcacacagctctatc 480
DB 155393 TTGCTGAGCATGTGATAGTATGATGACCAATCTCTGGGTGACACACACTCTTATTC 155334
QY 481 atgagctgtactctctgtgtggccaacacacatgaacctctctctgtgagataccacca 540
DB 155333 ATGAGCTGTACTCTCTGTGTGGCCAACACCATGACCACTCTCTCTGTGATGATACCCCA 155274
QY 541 ttgctgcttctcctgtgtagccctgtgaagatcaatgaagttggtgtgtgtgctgac 600
DB 155273 TTGCTGCTTCTCTCTCTTACCCCTGTAAGATCAATGAGTGTATGCTGTATGCTGTAT 155214
QY 601 attaccctgacataggagacattatctctacactgcatctctctatgcttcatctgt 660
DB 155213 ATTACCTGACATAGGAGGACTTATCTTACCGCATCTCTATGATGATTTATCATATGTT 155154
QY 661 gctattctcgtatccgacagtagaagcaagaagagagcctctcacaatgctcatct 720
DB 155153 GCTATTCTCCGTATCCGCACAGTAGAAGGCAAGAGAGGCTTCTTCAACATGCTCATCT 155094
QY 721 catctcaagctgtgagacccttactatctctctgtaactctacacatctatccgcctgtc 780
DB 155093 CATCTCAAGCTGTGAGACCTTACTATTTCTGTAATCTACACCTATATCCGCTGCT 155034
QY 781 tccagctatacatgtgaagaagacagagtgtagctcactcactatctctgtgactccc 840
DB 155033 TCCACCTATACATTTGAAAGAGACAAAGGTGTACTGACTCTTACTCTTTGTGACGCC 154974
QY 841 acattaacccagatgtgtgaagcttcacgaatagggagatgacgaagaagaatgaag 900
DB 154973 ACATTAACCCGATGTGTAGACGTTCCAGATAGGAGATGACGACGAGAAATTAGGAA 154914
QY 901 gtgttgcatcttctgaaacactag 924
DB 154913 GTGTTGCACTTTCTGAAACACTAG 154890
-----
RESULT 2
AC026038 138591 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-109C14, WORKING DRAFT
DEFINITION
SEQUENCE, 22 unordered pieces.
AC026038
AC026038.4 GI:9958133
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homini; Homo.
1 (bases 1 to 138591)
REFERENCE
1. The sequence of Homo sapiens clone
Waterson, R.H.
Unpublished
2 (bases 1 to 138591)
Waterson, R.H.
Direct Submission
Submitted (19-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:8567954.
-----
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information
Center project name: H.NH0109C14
----- Summary Statistics
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads

```

Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 129539 bases at least Q40
Consensus quality: 131919 bases at least Q30
Consensus quality: 133309 bases at least Q20
Insert size: 140000; agarose-fp
Insert size: 136491; sum-of-contigs
Quality coverage: 3.88 in Q20 bases; agarose-fp
Quality coverage: 4.04 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 5221: contig of 5221 bp in length
* 5222 5321: gap of unknown length
* 5322 9571: contig of 4250 bp in length
* 9572 9671: gap of unknown length
* 9672 13673: contig of 4002 bp in length
* 13674 13773: gap of unknown length
* 13774 18721: contig of 4948 bp in length
* 18722 18821: gap of unknown length
* 18822 24023: contig of 5202 bp in length
* 24024 24123: gap of unknown length
* 24124 29088: contig of 4865 bp in length
* 29089 34171: contig of 5083 bp in length
* 34172 34271: gap of unknown length
* 34272 40989: contig of 6718 bp in length
* 40990 41089: gap of unknown length
* 41090 48353: contig of 7266 bp in length
* 48356 48455: gap of unknown length
* 48456 59366: contig of 10911 bp in length
* 59367 61313: contig of 1847 bp in length
* 61314 61413: gap of unknown length
* 61414 71241: contig of 9828 bp in length
* 71242 80867: contig of 9526 bp in length
* 80868 80967: gap of unknown length
* 80968 99688: contig of 18721 bp in length
* 99689 99789: gap of unknown length
* 99790 122864: contig of 23076 bp in length
* 122865 122964: gap of unknown length
* 122965 124804: contig of 1840 bp in length
* 124805 124904: gap of unknown length
* 124905 126620: contig of 1716 bp in length
* 126621 126720: gap of unknown length
* 126721 128799: contig of 2079 bp in length
* 128800 128899: gap of unknown length
* 129000 130561: contig of 1662 bp in length
* 130562 130661: gap of unknown length
* 130662 132547: contig of 1886 bp in length
* 132548 132647: gap of unknown length
* 132648 135428: contig of 2781 bp in length
* 135429 135528: gap of unknown length
* 135529 138591: contig of 3063 bp in length.
```

FEATURES
source 1..138591
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-109C14"

misc_feature

/note="assembly_name:Contig10"

clone_end=:77

misc_feature

/note="assembly_name:Contig11"

```
misc_feature 9672..13673  
/note="assembly_name:Contig12"  
misc_feature 13774..18721  
/note="assembly_name:Contig13"  
misc_feature 18822..24023  
/note="assembly_name:Contig14"  
misc_feature 24124..28988  
/note="assembly_name:Contig15"  
misc_feature 29089..34171  
/note="assembly_name:Contig16"  
misc_feature 34272..40989  
/note="assembly_name:Contig17"  
misc_feature 41090..48355  
/note="assembly_name:Contig18"  
misc_feature 48456..59366  
/note="assembly_name:Contig19"  
misc_feature 59467..61313  
/note="assembly_name:Contig2"  
misc_feature 61414..71241  
/note="assembly_name:Contig20"  
misc_feature 71342..80867  
/note="assembly_name:Contig21"  
misc_feature 80968..99688  
/note="assembly_name:Contig22"  
misc_feature 99789..122864  
/note="assembly_name:Contig23  
clone_end:576  
vector_side:left"  
misc_feature 122965..124804  
/note="assembly_name:Contig3"  
misc_feature 124905..126620  
/note="assembly_name:Contig4"  
misc_feature 126721..128799  
/note="assembly_name:Contig5"  
misc_feature 128900..130561  
/note="assembly_name:Contig6"  
misc_feature 130662..132547  
/note="assembly_name:Contig7"  
misc_feature 132648..135428  
/note="assembly_name:Contig8"  
misc_feature 135529..138591  
/note="assembly_name:Contig9"
```

BASE COUNT 44211 a 25199 c 25318 g 41754 t 2109 others
ORIGIN

Query Match 99.7% Score 920.8; DB 2; Length 138591;
Best Local Similarity 99.8%; Pred. No. 3.4e-251;
Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 1 atgaatcacagtggtgaactgaattcaattatcttgggagcttcaaaaagctgaactc 60  
DB 86786 ATGAATCACAGGCTGTGTAATCTAGTTCAATTCTGGGCTTACCAAAAAGCTGAATC 86845  
OY 61 cagggaattatctctctcttcttctcattgtctatcttgtgcttcttcggcaaatg 120  
DB 86846 CAGGGAATTATCTCTCTCTTTTCTCAATGCTATCTTGTGGCTTTCTCGCAACATG 86905  
OY 121 cccatcatcatcttggcaaaatctatagcaaacctctgcatagcgccatgatatcttcc 180  
DB 86906 CTTCAATCATCATTCGCCAAATCTATTAACAACCTTGCAATACGCCCATATATGTTTCC 86965  
OY 181 ctgacactgagctggtgtgacatactgcacaaacatataccgaaagtctg99g 240  
DB 86966 CTGACACTGGCTGTGTGTGACATCTGCACAAAGAGATATACCGAATGCTG66g 87025  
OY 241 accatgtctaaatctcagaataatcattcatatctgagagctgcatgtccagctctctg 300  
DB 87026 ACCATGCTAAATCTCAGAAATATCATTCATATGAGAGGTGATGCCAGCTCTTCTTG 87085  
OY 301 ttcaatgtctctctgagagctgagatgattctctcaccacacatgctctatgacgcctat 360  
DB 87086 TTCAATGTCTCTCTGAGAGCTGAGATGTTCTTTCACCAACCATGCTATGACGCTAT 87145
```

OY	361	gtggccatttgttccctccttcattacagatcatataagnaccaccaatgtygtgtagcc	420
Db	87146	GTGGCCATTGTTTTCCTCCTCATTAACAGTAGACTGTATGAACCAACCATTATGTGTGTAGCC	87205
OY	421	ttagcccaacgatgcataatgacctatgcaatccaaatlcccttgtygtgcacaagaacttatc	480
Db	87206	TTGCTCACAGATGTCATGTGGCATATGTGCATTCACCAATTCTCGGTGACACAGCTCTTATC	87265
OY	481	atgaagtgtgaactttctgtggtgccaaaaccatatgaaccactctctctgtgagaatccccca	540
Db	87266	ATGAGGTGACATTTCTGTGGGCCAAAACCAATGACCACTTCTCTGTGAATATCCCCCA	87325
OY	541	ttagcgagcttctcctctgaagcccctaagaatcaatgagtgatgtygtatgtgtgta	600
Db	87326	TTCGTGCGCTTTCCTCCTGTAGCCCTGTAAAGAACATGAGGTGATGTATGTGCTGAT	87385
Db	601	attaaccctggccatagggaacttatcttaccctgcactcctctaagtgttatatgtt	660
Db	87386	ATTACCCCTTGCCATATGGGGACATTTATCTTACTCTCATCTCTATGGATTTATCATTTGT	87445
OY	661	gctattctccgfatccgcacagtaagaagcaagaagagcccttccaacatgctcatct	720
Db	87446	GCTATTCTCCGATCCGCACTAAGAAGGCAGAGAGAGCCCTTCCAACAATGCTCATCT	87505
OY	721	catctcaagtygtgtgaeccttactatcttctcctgtaatctaacactatataccgcctct	780
Db	87506	CATCTCACAGTGTGATGCCCTTTACTATTTCTCTGTATCTTACACCTATATCCGCCCTGCT	87565
OY	781	tccagctatacatattgaagaagacaagtygtgtgtagctgcactatatactcttgtgaatccc	840
Db	87566	TCCAGCTATATACATTGAAGAAGACAAAGTGTGATGTCACCTCATATCTGTGTGACTCCC	87625
OY	841	acaattaaaccgatgtygtacagcttcccgaataaggyagaatccaagccaagaattiggaag	900
Db	87626	ACATTTAAACCCGATGTGTATACACCTTCCAGAAATGGAATGGAGATGCAAGCAGCAATTGAGAAG	87685
OY	901	gtgttgcatctctgaaacactag 924 	
Db	87686	GTGTTTGCAATTTCTGAAACACTAG 87709 	
RESULT	3		
AL357039		190889 bp DNA linear HTG 19-SEP-2001	
LOCUS		Homo sapiens chromosome 1 clone RP11-63AB7, *** SEQUENCING IN	
DEFINITION		PROGRESS ***, 18 unordered pieces.	
ACCESSION	AL357039		
VERSION	AL357039.5	GI:9797648	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
JOURNAL		Plumb.B. Direct Submission Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Aug 12, 2000 this sequence version replaced gi:921404.	
COMMENT		Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk Project Information Center project name: BA634B7 ----- Summary Statistics ----- Assembly program: XGAP4; Version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 183408 bases at least Q40	

Consensus quality: 186557 bases at least Q30
Consensus quality: 188087 bases at least Q20
Insert size: 189189; sum-of-contigs
Insert size: 188721; 6.6% error; agarose-fp
Quality coverage: 4.17x in Q20 bases; sum-of-contigs quality
coverage: 4.30x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 colligs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the colligs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* De pressed: 12042: contig of 12042 bp in length
* 1
* 12043 12142: gap of 100 bp
* 12143 12143 30262: contig of 18120 bp in length
* 30263 30362: gap of 100 bp
* 30363 41425: contig of 11063 bp in length
* 41426 41525: gap of 100 bp
* 41526 48351: contig of 6826 bp in length
* 48352 48451: gap of 100 bp
* 48452 57808: contig of 9357 bp in length
* 57809 57908: gap of 100 bp
* 57909 68643: contig of 10735 bp in length
* 68644 68743: gap of 100 bp
* 68744 75536: contig of 6793 bp in length
* 75537 75636: gap of 100 bp
* 75637 84190: contig of 8554 bp in length
* 84191 84280: gap of 100 bp
* 84281 89038: contig of 4749 bp in length
* 89040 89139: gap of 100 bp
* 89140 94516: contig of 3377 bp in length
* 94517 94616: gap of 100 bp
* 94617 105920: contig of 11304 bp in length
* 105921 106020: gap of 100 bp
* 106021 133564: contig of 27544 bp in length
* 133565 133664: gap of 100 bp
* 133665 137940: contig of 4276 bp in length
* 137941 138040: gap of 100 bp
* 138041 141386: contig of 3346 bp in length
* 141387 141486: gap of 100 bp
* 141487 143513: contig of 2027 bp in length
* 143514 143613: gap of 100 bp
* 143614 159175: contig of 15562 bp in length
* 159176 159275: gap of 100 bp
* 159276 174409: contig of 15134 bp in length
* 174510 174509: gap of 100 bp
* 174510 190889: contig of 16380 bp in length

```

```
source
1. .190889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-634B7"
/clone_id="RP11-634B7"
1. .12042
/note="assembly_fragment:01421
fragment_chain:1
clone_end:77
vector_side:left"
12143. .3062
/note="assembly_fragment:00925
fragment_chain:1"
30363. .41425
/note="assembly_fragment:00451
fragment_chain:2"
41536. .48351
/note="assembly_fragment:00297
fragment_chain:2"
48432. .57808
/note="assembly_fragment:00082
```



```

misc_feature      fragment_chain:2"
                    57909..68643
                    /note="assembly_fragment:01434
                    fragment_chain:2"
misc_feature      68744..75536
                    /note="assembly_fragment:01941
                    fragment_chain:2"
misc_feature      75637..84190
                    /note="assembly_fragment:00841
                    fragment_chain:2"
misc_feature      84291..89039
                    /note="assembly_fragment:00517
                    fragment_chain:3"
misc_feature      89140..94516
                    /note="assembly_fragment:00194
                    fragment_chain:3"
misc_feature      94617..105920
                    /note="assembly_fragment:00685
                    fragment_chain:4"
misc_feature      106021..133564
                    /note="assembly_fragment:00244
                    fragment_chain:4"
misc_feature      133665..137940
                    /note="assembly_fragment:00993
                    fragment_chain:5"
misc_feature      138041..141386
                    /note="assembly_fragment:00049
                    fragment_chain:5"
misc_feature      141487..143513
                    /note="assembly_fragment:01137"
                    fragment_chain:5"
misc_feature      143614..159175
                    /note="assembly_fragment:01189"
                    fragment_chain:5"
misc_feature      159276..174409
                    /note="assembly_fragment:01199"
                    fragment_chain:5"
misc_feature      174510..190889
                    /note="assembly_fragment:01519"
                    fragment_chain:5"
BASE COUNT      60637 a 34972 c 35649 g 57919 t 1712 others
ORIGIN

```

```

Query Match      99.7%; Score 920.8; DB 2; Length 190889;
Best Local Similarity 99.8%; Pred. No. 3.5e-251;
Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgaatcacagcgtgtgaactgaatcattctatctggcgctacacaaagcctgaacc 60
    |||||||
DB 116178 ATGAATCACAGCGTGTGAATCAGTTCAATTTCTGGGCTCACCAAAAGCCTGAATC 116237

QY 61 cagggaattatctctctctctctctctctctctctctctctctctctctctctctct 120
    |||||||
DB 116238 CAGGGAATTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 116297

QY 121 ctcatcatcatctgcacaaatctatagcaacacctgacacgacacgacacgacacgac 180
    |||||||
DB 116298 CTCATCATCATCTGCACAAATCTATACACACACCTTGCAATCCCATGTATCTTTCT 116357

QY 181 ctgacacacgagcgtgtgtgacacacacacacacacacacacacacacacacacacac 240
    |||||||
DB 116358 CTGACACTGCGCTGTGTGACATCTGCACACACACGATCATCCGAAATGCTGGGG 116417

QY 241 accatgctacaacacgaacaaatcacatctcatatgacgctgacacgctctctctctg 300
    |||||||
DB 116418 ACCATGCTACAACGAAATTCATTCATTCATTCGACGCTGATGCCAGCTCTTCTTG 116477

QY 301 ttcaaatgtctctgtgagcgtgagatggtctctctcaacacacacacacacacacacacac 360
    |||||||
DB 116478 TTCAATGTCTCTGTGAGCTGAGATGTTCTCTCACACACCATGCGCTATGACCGTAT 116537

QY 361 gtggacattgtctccctctcatcaagctactatgaacacacacacacacacacacacac 420
    |||||||
DB 116538 GTGGCAATTTGTTCTCCCTCTTATTAAGTACTGTATGAACACCATATGTTGTAGGCG 116597

QY 421 ttgctcaagatgltatgctatgcaatlcacaaatctctggtgacacacacacacacacac 480

```

```

DB 116598 TTGCTACAGATGTCATGCTATTCAGTACACCAATTCCTGGGTGCACACAGCTTTATTC 116657
    |||||||
QY 481 atgaggttgaacttctctgtgtggccaacacacacacacacacacacacacacacacacac 540
    |||||||
DB 116658 ATGAGGTGACTTCTGTGTGGGCCAAACACCATTTGACCACTTCTGTGTAGATACCCCA 116717

QY 541 ttgctgcttctctctgtgtgacccctgtgaacacacacacacacacacacacacacacacac 600
    |||||||
DB 116718 TTGCTGCTTGTGCTGTGTGACCCCTGTGAATCAATGAAGTGTATGTGTGTGTGTGT 116777

QY 601 ataaccttgcccaatgaaggaacttatctctacatgacatccatgacacacacacacacac 660
    |||||||
DB 116778 ATTACCTTGCCCATGAGGACATTTATTTCTTACTGCATCTCTCATGTTTATCATTTGTT 116837

QY 661 gctatctccglatccgcacagtaagaagcacaaggaagcctctcaacacacacacacacac 720
    |||||||
DB 116838 GCTATTCCTCCGATATCCGACAGATAGAAAGGCAAGAGAGAGGCTTCTCAACATGCTCATCT 116897

QY 721 catctcacagtggtgaccccttactatctctctgtatcttaacacacacacacacacacac 780
    |||||||
DB 116898 CATCTCACAGTGTGACCCCTTACTATTTCTCTGTATATACACCTATATCCGCTGCT 116957

QY 781 tcaagctacacatttgaagagacaagtggtgagctgcacactatactctgtgactccc 840
    |||||||
DB 116958 TCCAGCTATACATTTCAAAAGACACAGAGTGATGCTGCACTATATCTTTGTGACTGCC 117017

QY 841 acaatlaaacccagatggtgtacagctctccagaaataggaagacagcagaagaaataggaag 900
    |||||||
DB 117018 ACATTTAAACCCGATGTGTACAGCTTCCAGAAATAGGAGATGACAGCAGAAATTAGAAG 117077

QY 901 gtgtgtgcatcttcgaacacactag 924
    |||||||
DB 117078 GTGTTTGCAATTTCTGAACACGTAG 117101

```

```

RESULT 4
AX241862          921 bp      DNA          linear      PAT 26-SEP-2001
LOCUS             AX241862          Sequence 610 from Patent WO0127158.
DEFINITION        AX241862          AX241862
ACCESSION         AX241862          GI:15798737
VERSION           AX241862.1
KEYWORDS          synthetic construct.
SOURCE            synthetic construct.
ORGANISM          artificial sequence.
REFERENCE         1 (bases 1 to 921)
AUTHORS           Bellenson,J., Smith,D., Lancel,D., Glusman,G., Fuchs,T. and
                  Yanai,I.
TITLE             Olfactory receptor sequences
JOURNAL           Patent: WO 0127158-A 610 19-APR-2001.
FEATURES          Location/Qualifiers
                   1..921
                   /organism="synthetic construct"
                   /db_xref="taxon:32630"
                   /note="(H389459 nucleotide)"
BASE COUNT       218 a 235 c 183 g 285 t
ORIGIN

```

```

Query Match      99.3%; Score 917.8; DB 6; Length 921;
Best Local Similarity 99.8%; Pred. No. 1.5e-250;
Matches 919; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgaatcacagcgtgtgaactgaatcattctctgtggcctcacaacaaagcctgaacc 60
    |||||||
DB 1 ATGAATCACAGCGTGTGAATCAGTTCAATTTCTGGGCTCACCAAAAGCCTGAATC 60

QY 61 cagggaattatctctctctctctctctctctctctctctctctctctctctctctctct 120
    |||||||
DB 61 CAGGGAATTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120

```

OY 121 ctcacatcatcgtcccaaaatctatagcaaaccttgcatacagcccatgatatgttctc 180
 Db 121 ctcacatcatcgtcccaaaatctatagcaaaccttgcatacagcccatgatatgttctc 180
 OY 181 ctgacacgtgctgttctggaacatcctgcacaaacgaatacagaagatgctggg 240
 Db 181 ctgacacgtgctgttctggaacatcctgcacaaacgaatacagaagatgctggg 240
 OY 241 accatgctaacatcaaaaataacattctatctgagctgctgctcctcctc 300
 Db 241 accatgctaacatcaaaaataacattctatctgagctgctgctcctcctc 300
 OY 301 ttacacatgctcctggagctgagatgctcctcctcacaacattgagctcctc 360
 Db 301 ttacacatgctcctggagctgagatgctcctcctcacaacattgagctcctc 360
 OY 361 gtggccattgttccctcctcctcctcctcctcctcctcctcctcctcctc 420
 Db 361 gtggccattgttccctcctcctcctcctcctcctcctcctcctcctcctc 420
 OY 421 ttgttcagcagctgctcagctgctcctcctcctcctcctcctcctcctc 480
 Db 421 ttgttcagcagctgctcagctgctcctcctcctcctcctcctcctcctc 480
 OY 481 atgaggttgcacttctgttggccaacacacattgacacattcctcctcctc 540
 Db 481 atgaggttgcacttctgttggccaacacacattgacacattcctcctcctc 540
 OY 541 ttgttcagcttctgtcctcctcctcctcctcctcctcctcctcctcctc 600
 Db 541 ttgttcagcttctgtcctcctcctcctcctcctcctcctcctcctcctc 600
 OY 601 attaccctgcagcaagggagcttctcctcctcctcctcctcctcctcctc 660
 Db 601 attaccctgcagcaagggagcttctcctcctcctcctcctcctcctcctc 660
 OY 661 gctattccctcctcctcctcctcctcctcctcctcctcctcctcctcctc 720
 Db 661 gctattccctcctcctcctcctcctcctcctcctcctcctcctcctcctc 720
 OY 721 catcctcagctgctgacacagctgaggaaggaaggaaggaaggaaggaag 780
 Db 721 catcctcagctgctgacacagctgaggaaggaaggaaggaaggaaggaag 780
 OY 781 tccagctatacatttgaaggaaggaaggaaggaaggaaggaaggaaggaag 840
 Db 781 tccagctatacatttgaaggaaggaaggaaggaaggaaggaaggaaggaag 840
 OY 841 acattaaacccgagtgctgacagcttccagaaataggaaggaaggaaggaag 900
 Db 841 acattaaacccgagtgctgacagcttccagaaataggaaggaaggaaggaag 900
 OY 901 gttgttgcatttctcgaacac 921
 Db 901 gttgttgcatttctcgaacac 921
 RESULT 5
 AX350633 798 bp DNA linear PAT 06-FEB-2002
 LOCUS AX350633 Sequence 11 from Patent WO01771177.
 DEFINITION AX350633
 ACCESSION AX350633
 VERSION AX350633.1 GI:18616201
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Shenoy,S., Gangolli,E.A., Rastelli,L., Smithson,G., Padigaru,M.,
 Verneet,C.A., Wolenc,A.R., Casman,S.J., Tcheney,V.T.,
 Szekeres,E.S., Gorse,W., Alsbrook,J.P. and Burgess,C.E.
 TITLE Novel spec-proteins and nucleic acids encoding same

JOURNAL Patent: WO 0177177-A 11 18-OCT-2001;
 FEATURES Curagen Corporation (US)
 source Location/Qualifiers
 1. 798
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 191 a 202 c 164 g 241 t
 ORIGIN
 Query Match 84.3%; Score 778.8; DB 6; Length 798;
 Best Local Similarity 99.7%; Pred. No. 6.2e-211;
 Matches 780; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 143 atagaacaccttgcatacgcctcctcctcctcctcctcctcctcctcctc 202
 Db 143 atagaacaccttgcatacgcctcctcctcctcctcctcctcctcctcctc 202
 OY 203 tcatctgacaacaagaacatcacaacgaagatgctgggacatgctcaacataa 262
 Db 203 tcatctgacaacaagaacatcacaacgaagatgctgggacatgctcaacataa 262
 OY 61 tcatctgacaacaagaacatcacaacgaagatgctgggacatgctcaacataa 120
 Db 61 tcatctgacaacaagaacatcacaacgaagatgctgggacatgctcaacataa 120
 OY 263 ccatttcatacagcagcagcagcagcagcagcagcagcagcagcagcagc 322
 Db 263 ccatttcatacagcagcagcagcagcagcagcagcagcagcagcagcagc 322
 OY 121 ccatctcattatgcagcagcagcagcagcagcagcagcagcagcagcagc 180
 Db 121 ccatctcattatgcagcagcagcagcagcagcagcagcagcagcagcagc 180
 OY 323 agatgctcctcctcacaacacagcagcagcagcagcagcagcagcagcagc 382
 Db 323 agatgctcctcctcacaacacagcagcagcagcagcagcagcagcagcagc 382
 OY 181 agatgctcctcctcacaacacagcagcagcagcagcagcagcagcagcagc 240
 Db 181 agatgctcctcctcacaacacagcagcagcagcagcagcagcagcagcagc 240
 OY 383 attacagctatctatgaacacacatagctgctgctgctgctgctgctgct 442
 Db 383 attacagctatctatgaacacacatagctgctgctgctgctgctgctgct 442
 OY 241 attacagctatctatgaacacacatagctgctgctgctgctgctgctgct 300
 Db 241 attacagctatctatgaacacacatagctgctgctgctgctgctgctgct 300
 OY 443 ttgcagcacaac 502
 Db 443 ttgcagcacaac 502
 OY 301 ttgcagcacaac 360
 Db 301 ttgcagcacaac 360
 OY 503 caaacacacattgcacacacacacacacacacacacacacacacacacacac 562
 Db 503 caaacacacattgcacacacacacacacacacacacacacacacacacacac 562
 OY 361 caaacacacattgcacacacacacacacacacacacacacacacacacacac 420
 Db 361 caaacacacattgcacacacacacacacacacacacacacacacacacacac 420
 OY 563 ctgtaagaacatcagagctgctgctgctgctgctgctgctgctgctgctgct 622
 Db 563 ctgtaagaacatcagagctgctgctgctgctgctgctgctgctgctgctgct 622
 OY 421 ctgtaagaacatcagagctgctgctgctgctgctgctgctgctgctgctgct 480
 Db 421 ctgtaagaacatcagagctgctgctgctgctgctgctgctgctgctgctgct 480
 OY 623 ttaattcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 682
 Db 623 ttaattcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 682
 OY 481 ttattcttaccctcctcctcctcctcctcctcctcctcctcctcctcctcct 540
 Db 481 ttattcttaccctcctcctcctcctcctcctcctcctcctcctcctcctcct 540
 OY 683 tagaagcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 742
 Db 683 tagaagcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 742
 OY 541 tagaagcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 600
 Db 541 tagaagcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 600
 OY 743 actatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 802
 Db 743 actatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 802
 OY 601 actatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 660
 Db 601 actatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 660
 OY 803 acaaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 862
 Db 803 acaaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 862
 OY 863 gcttcagaagaataggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 922
 Db 863 gcttcagaagaataggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 922
 OY 721 gcttcagaagaataggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 780
 Db 721 gcttcagaagaataggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 780
 OY 923 ag 924
 Db 923 ag 924
 OY 781 ag 782
 Db 781 ag 782
 RESULT 6
 AC099601/c

LOCUS	AC099601	22166 bp	DNA	linear	HTG 16-NOV-2001
DEFINITION	Mus musculus clone RP23-240K1 WORKING DRAFT SEQUENCE, 3 unordered pieces.				
ACCESSION	AC099601				
VERSION	AC099601.1 GI:16946330				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
AUTHORS	Eunhyota Metaxoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Mus musculus, clone RP23-240K1				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 22166)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Baran,N., Bastien,V., Boguslavskiy,L., Boukhalter,B., Brown,A., Camarata,R., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Harford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kamat,A., Karatsis,A., Kells,C., Larocque,K., MacIazens,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,K., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McNetters,R., Meldrum,J., Menues,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhand,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Ribback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Tigillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,M.J., Young,G., Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.				
COMMENT	Direct Submission Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA All repeats were identified using RepeatMasker: Smit,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
TITLE	Genome Center				
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: WIBR				
	Web site: http://www-seq.wi.mit.edu				
	Contact: sequence_submissions@genome.wi.mit.edu				
	----- Project Information -----				
	Center project name: LI7240				
	Center clone name: 240_K_1				
	----- Summary Statistics -----				
	Sequencing vector: Plasmid; n/a; 100% of reads				
	Chemistry: Dye-terminator Big Dye; 100% of reads				
	Assembly program: Phrap; version 0.960731				
	Consensus quality: 220831 bases at least Q40				
	Consensus quality: 221147 bases at least Q30				
	Consensus quality: 221206 bases at least Q20				
	Insert size: 230000; agarose-fp				
	Insert size: 221266; sum-of-contigs				
	Quality coverage: 11.4 in Q20 bases; agarose-fp				
	Quality coverage: 11.9 in Q20 bases; sum-of-contigs				
	----- NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.				
*	* 1 9911: contig of 9911 bp in length				
*	* 9912 10011: gap of 100 bp				

FEATURES	source	10012	74781	contig of 64770 bp in length
*		74782	74881	gap of 100 bp
*		74882	221466	contig of 146585 bp in length.
	Location/Qualifiers			
	1..221466			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="RP23-240K1"			
	/clone_id="RPC1-23 Female Mouse BAC"			
	1..9911			
	/note="assembly_fragment"			
	clone_end:SP6			
	vector_side:left"			
	10012..74781			
	/note="assembly_fragment"			
	74882..221466			
	/note="assembly_fragment"			
	clone_end:T7			
	vector_side:right"			
	BASE COUNT	71567 a	39922 c	39522 g
	ORIGIN		70247 t	208 others
	Query Match	74.4%	Score 687.8;	DB 2: Length 221466;
	Best Local Similarity	84.1%;	Pred. No. 9.1e-185;	
	Matches	776;	Conservative	0; Mismatches 147; Indels 0; Gaps
QY	1	atgaatcacagcgtgtgtactgagttcatatctctggcctcaccaaaagcctgaac	60	
Db	193321	ATGAACTTCACACATGTTAGTAGAGTTATGATCTGGGACTTACTCAAAAGCTGAACTT	193262	
QY	61	cagggaatatactctccctctttcttcctcatgtgtcattctgtgtcttcttcggcaacatg	120	
Db	193261	CAGGGAATACCTTTCATTGTTTCTCTTTTATTCACCTTGGCTCTACCTGGTAATATG	193202	
QY	121	ctcatcatcatcttgccaaaactctatagaacaacctgtgatagcccatgtatgtttccctt	180	
Db	193201	CTAATGTGTGTGGCATATATCTATACACACCCTTGCACACACCACCATGTATATTCCTT	193142	
QY	181	ctgacactgctgtgtgtgacatcatctgacacaacaagcatcacaaccgaagatgctggg	240	
Db	193141	CTGGCTTGTCGTGGGTGGAGCATATATGTCGACCTCAAGCATCATACCAAAATGTTGGGA	193082	
QY	241	accatgtctacaatacaagaataatcacattcatatgacagctgcatgtgtccagctctctg	300	
Db	193081	ACTATGTTAACATCAAAAAAATTCAATTTCAATTTGGGGTTGCAATGCCACGCTCTCTT	193022	
QY	301	ttcacatgtctctcggagagctggaagtgtctcttccacacacagagcctatgaacctat	360	
Db	193021	TTCACATGGCTCTGGGGGCTGAGATGGTCTTTTACTTCMAATGGCTTATGACCGCTAT	192962	
QY	361	gtggacatttgttccctctctcatatcacagtaacttlatgaaccaacatagtgtgagcc	420	
Db	192961	GTGGCATTTTGCTTCCCACTTCGGTATAGTACTATTTATGAACACATATATGATGTAGGC	192902	
QY	421	ttgtctcaagcatgtaatgtaatgtcatgtgcagtcacaaattccctgggtgacacagctctatc	480	
Db	192901	TTTGCTTACCATTTGCATGGCATTTGCTGTAAACCAATTCTGGGGTGCACACTGGTCTCAT	192842	
QY	481	atgaggttgacttctctgtgggccaacaacacatgtgacacattctctctgtgtgataacccca	540	
Db	192841	CTGAAGCTGACTTTCTGTGGGCCCAAAATATTAATTCACCACTTCTCTGTGAATAACCCCG	192782	
QY	541	tttgcgtgtgtgtctgtgagccctgtgaagaatcaatgaagtgtatgtgtgtgtgtgtgtgt	600	
Db	192781	CTGCTGTGTTTGCTGTAGTCCGTGTAGAGTCAATAGGTATGGTGTGTGTGTGTGTGTGTGT	192722	
QY	601	attaccctggcatagagggaattatcttactacgtgcatccccaatgtttatcatatgtt	660	
Db	192721	ATCAACCGCGGTGTGGGAGATTATACCTAAACGTGCAATTCGATGTATATTAATTTGCT	192662	
QY	661	gctattctcgatctcgacacagtagaagcgaagaaggcctcttccacaatgctcatct	720	

	Db	192661	GCTATTTCGCCATCCGACACAGCAAGAACGAAGAAGCGCTTCTGTACTGCTCAGC	192602
Oy	721	cattccacagatgctgtgaacccttaactatctcctcglaatcatcacacctataccgccctgc	780	
Db	192601	CACCTCAGTGAGTGATGGTCCCTTTACTATTCTCTCGTTACTTCTACCTATATTCGCAGCTGCA	192542	
Oy	781	tccagccttaacttggaaagaagacaaggctgtgaagtgcactctaactcttgtaactccc	840	
Db	192541	TCCACTCTATACCTTTGATAAAGACAAGGTGGTAGTGCACATATATACCTTGTAACCTCT	192482	
Oy	841	acaatcaaacccgatgtgtgttacagcttcacagatataggsagatgcaggaagaattaagag	900	
Db	192481	ACATTGAACCCCAATATGTAATATAGCTTTCAGGAACAAGAGATGCAAGTATTAAGAA	192422	
Oy	901	gtgtttgcattcttgaacacta	923	
Db	192421	GTAATTGCAATTTTTTAAGGTTA	192399	
	RESUTJ.F.			
	7			
	LOCUS	AC094718	160483 bp DNA linear HTG 20-DEC-2001	
	DEFINITION	Rattus norvegicus clone CH230-5M10,	*** SEQUENCING IN PROGRESS ***	
		65 unordered pieces.		
	ACCESSION	AC094718		
	VERSION	AC094718.2 GI:17941497		
	KEYWORDS	HTG; PHASEL.		
	SOURCE	Norway rat.		
	ORGANISM	Rattus norvegicus		
		Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi;		
		Eumetazoa; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Murinae;		
		Rattus.		
		1 (bases 1 to 160483)		
		Muzny,D.M., Adams,C.C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,		
		Alshrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,U.,		
		Benton,J., Binage,K., Blankenburg,K., Bonin,D., Boucek,J.,		
		Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,		
		Burch,P., Burelt,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,		
		Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,		
		Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,		
		Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,		
		Davy-Carroll,L., Deckerich,D.A., Delaney,K.R., Delgado,O.,		
		Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,		
		Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,		
		Elhaj,C., Escotto,M., Falls,T., Farguto,D., Flagg,N., Ford,J.,		
		Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,		
		Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Guarnatine,P., Hale,S.,		
		Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,C.,		
		Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,		
		Hollins,B., Homsí,F., Howard,S., Huber,J., Huliy,S., Hume,J.,		
		Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivett,S.,		
		Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,		
		Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,		
		Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,		
		Louisegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R.,		
		Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,		
		Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,		
		Mei,C., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,		
		Morgan,M., Morris,S., Moser,M., Neale,D., Newton,J., Newton,N.,		
		Morgen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwu,S.,		
		Ogunu,M., Okunonu,G., Oraguwe,N., Oviledo,R., Pace,A., Payton,B.,		
		Pearcy,J., Peery,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,		
		Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuchakan,I., Rolfe,M.,		
		Ritz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshcharil,N.,		
		Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,		
		Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,		
		Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,		
		Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,		
		Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,		
		Wellington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,		
		Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,		
		Weinstock,G. and gibbs,R.		
		Direct Submission		
	TITLE			

Unpublished
2 (bases 1 to 160483)
Worley K.C.
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (15-Sep-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624554.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GRP
Center clone name: CH230-SM10

Summary Statistics
Assembly program: Phrap; version 0.990329p1r1st call to
findPrapl1st

Consensus quality: 127493 bases at least Q40
Consensus quality: 137458 bases at least Q30
Consensus quality: 144882 bases at least Q20
Estimated insert size: 129461; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
5754: contig of 5754 bp in length
*
* 5755 5854: gap of unknown length
*
* 5855 11490: contig of 5636 bp in length
*
* 11491 11590: gap of unknown length
*
* 11591 17216: contig of 5626 bp in length
*
* 17217 17316: gap of unknown length
*
* 17317 21754: contig of 4438 bp in length
*
* 21755 21854: gap of unknown length
*
* 21855 25351: contig of 3497 bp in length
*
* 25352 25451: gap of unknown length
*
* 25452 30651: contig of 5200 bp in length
*
* 30652 30751: gap of unknown length
*
* 30752 34861: contig of 4110 bp in length
*
* 34862 34961: gap of unknown length
*
* 34962 38499: contig of 3538 bp in length
*
* 38500 38599: gap of unknown length
*
* 38600 41784: contig of 3185 bp in length
*
* 41785 41884: gap of unknown length
*
* 41885 45873: contig of 3989 bp in length
*
* 45874 45973: gap of unknown length
*
* 45974 50224: contig of 4251 bp in length
*
* 50225 50324: gap of unknown length
*
* 50325 53266: contig of 2942 bp in length
*
* 53267 53367: gap of unknown length
*
* 53368 56430: contig of 3064 bp in length
*
* 56431 56530: gap of unknown length
*
* 56531 59835: contig of 3305 bp in length
*
* 59836 59935: gap of unknown length
*
* 59936 62863: contig of 2928 bp in length
*
* 62864 62963: gap of unknown length
*
* 62964 65355: contig of 2392 bp in length
*
* 65356 65455: gap of unknown length
*
* 65456 68305: contig of 2850 bp in length
*
* 68306 68405: gap of unknown length
*
* 68406 70717: contig of 2312 bp in length
*
* 70718 73306: contig of 2489 bp in length
*
* 73307 73306: contig of 2489 bp in length

AY073900 926 bp DNA linear ROD 04-FEB-2002
 LOCUS AY073900 Mus musculus olfactory receptor MOR251-4p pseudogene, partial
 DEFINITION sequence.
 ACCESSION AY073900
 VERSION AY073900.1 GI:18481062
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 926)
 AUTHORS Zhang, X. and Firestein, S.
 TITLE The olfactory receptor gene superfamily of the mouse
 JOURNAL Nat. Neurosci. 5 (2), 124-133 (2002)
 PUBMED 11802173
 REFERENCE 2 (bases 1 to 926)
 AUTHORS Adams, M.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 FEATURES
 source Location/Qualifiers
 1..926
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="7"
 <1..>926
 /gene="olfactory receptor MOR251-4p"
 /note="potential coding region disrupted by one or more
 internal stop codons"
 /pseudo
 /evidence=not-experimental
 BASE COUNT 227 a 209 c 181 g 309 t
 ORIGIN
 Query Match 72.0%; Score 665.6; DB 10; Length 926;
 Best Local Similarity 84.1%; Pred. No. 1.1e-178;
 Matches 774; Conservative 0; Mismatches 144; Indels 2; Gaps 2;

Db 484 CTGAGCTGACTTCTGTGGCGGAATATATGACACATCTTCTGTGAATATACCCC 543
 Qy 539 cattgctggttgcctctgtgaagcctgaagatcaatgagtgatggtgctg 598
 Db 544 CGTGTGCTTGTGCTGCTGAGTCCGTGAAGCAATGAGTGAAGTGTATGCTGCG 603
 Qy 599 atattaccctggcattagggacttattcttactgcacatctctatggtttatcatg 658
 Db 604 ATATACACCTGGCTGTGGAGACTTACTCAACCTCAATTCCTATGATTTATTTG 663
 Qy 659 ttgtattcttcggtatccgcagtagaaggcaaggagggcctctcaactgcat 718
 Db 664 CTGCATATCTCCGATCCGACAAACAAAGCAAGAAAGGAGGAGGCTTCTACCTGCTCAT 723
 Qy 719 ctcaatcgaagtgtagaccccttactattctcctgttaactcaactatccgcctg 778
 Db 724 CGACCTCATGAGTGTGCTGCTTACTATCTCTGTTATCTACCATATATCCGACTG 783
 Qy 779 ctccagctatacatltyaaagagaagaagtgtagctgcacatcttactcttgactc 838
 Db 784 CATCCAGCTATACCTTGTATTAAGCAAGGTGTAGCTGCATATATCTTACTTAATCTC 843
 Qy 839 ccacattaaacccgtagtgtagacgctccagaataagagatgcaggcaggaaataga 898
 Db 844 CTACATGTGAACCAATATGATATATCTTCTCAGGAACAAAGATGACAGCAATTAAGA 903
 Qy 899 aggtgttgcatcttctgaa 918
 Db 904 AAGTATTTGCATTTTAA 923
 RESULT 9
 AX350675 663 bp DNA linear PAT 06-FEB-2002
 LOCUS AX350675 Sequence W00171717.
 DEFINITION AX350675
 ACCESSION AX350675
 VERSION AX350675.1 GI:18616221
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Shenoy, S., Gangolli, E.A., Rastelli, L., Smithson, G., Padigaru, M.,
 Vernet, C.A., Wolenc, A.R., Casman, S.J., Tcheney, V.T.,
 Szekeres, E.S., Gorsse, W., Alsobrook, J.P., and Burgess, C.E.
 TITLE Novel spcr proteins and nucleic acids encoding same
 JOURNAL Patient: WO 017717-A 53 18-OCT-2001;
 Curagen Corporation (US)
 FEATURES
 source Location/Qualifiers
 1..663
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 151 a 174 c 134 g 204 t
 ORIGIN
 Query Match 71.6%; Score 661.4; DB 6; Length 663;
 Best Local Similarity 99.8%; Pred. No. 1.7e-177;
 Matches 662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 367 atttcttccctctcattacagactactatataaaccacatattgtatagcttgc 426
DB 181 atttcttccctctcattacagactactatataaaccacatattgtatagcttgc 240
OY 427 agcatgtcatcagctatcagctacccaattctctgtgtgcacacagctcttcatgaag 486
DB 241 accatgtcatcagctatcagctacccaattctctgtgtgcacacagctcttcatgaag 300
OY 487 ttgactctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 546
DB 301 ttgactctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
OY 547 gcttctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 606
DB 361 gcttctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
OY 607 ctgtgcataaggagacttattcttaccctgcatctcctcattgtttatcatattgtctatt 666
DB 421 ctgtgcataaggagacttattcttaccctgcatctcctcattgtttatcatattgtctatt 480
OY 667 ctccgtatccgacagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 726
DB 481 ctccgtatccgacagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 540
OY 727 acaagtgtagacccttactatctctctgtatctacactatatacgcctctccagc 786
DB 541 acagtgtagacccttactatctctctgtatctacactatatacgcctctccagc 600
OY 787 tatactttgaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 846
DB 601 tatactttgaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 660
OY 847 aac 849
DB 661 AAC 663

RESULT 10
AC101272 59688 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-101J20, LOW-PASS SEQUENCE SAMPLING.
AC101272
AC101272.1 GI:17060047
VERSION HTG: HTGS_PHASE0.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 59688)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-101J20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 59688)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, J., Bouckgeert, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choe, P., Colangelo, M., Collins, S., Collins, S., Cook, A.,
Cooke, P., DeArnell, K., Dewar, K., Diaz, J., Dodge, S., Fato, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Gardner, L., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamarez, R., Landers, T., Lehoucq, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Melidrim, J.,
Menou, C., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norman, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

```

TITLE JOURNAL COMMENT

```

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L16346
Center clone name: 101_J_20
-----
* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
679: contig of 679 bp in length
680 779: gap of 100 bp
780 1489: contig of 710 bp in length
1490 1589: gap of 100 bp
1590 2263: contig of 674 bp in length
2264 2363: gap of 100 bp
2364 3054: contig of 691 bp in length
3055 3154: gap of 100 bp
3155 3839: contig of 685 bp in length
3840 3939: gap of 100 bp
3940 4630: contig of 691 bp in length
4631 4730: gap of 100 bp
4731 5401: contig of 671 bp in length
5402 5501: gap of 100 bp
5502 6164: contig of 663 bp in length
6165 6264: gap of 100 bp
6265 6949: contig of 685 bp in length
6950 7049: gap of 100 bp
7050 7728: contig of 679 bp in length
7729 7828: gap of 100 bp
7829 8516: contig of 688 bp in length
8517 8616: gap of 100 bp
8617 9325: contig of 709 bp in length
9326 9425: gap of 100 bp
9426 10127: contig of 702 bp in length
10128 10227: gap of 100 bp
10228 10924: contig of 697 bp in length
10925 11024: gap of 100 bp
11025 11712: contig of 688 bp in length
11713 11812: gap of 100 bp
11813 12481: contig of 669 bp in length
12482 12581: gap of 100 bp
12582 13254: contig of 673 bp in length
13255 13354: gap of 100 bp
13355 14033: contig of 679 bp in length
14034 14133: gap of 100 bp
14134 14816: contig of 683 bp in length
14817 14916: gap of 100 bp
14917 15597: contig of 661 bp in length
15598 15697: gap of 100 bp
15698 16388: contig of 691 bp in length
16389 16488: gap of 100 bp
16489 17234: contig of 746 bp in length
17235 17334: gap of 100 bp

```


Assembly program: Phrap; version 0.990329First call to findphraplist

Consensus quality: 96697 bases at least Q40
 Consensus quality: 104647 bases at least Q30
 Consensus quality: 112341 bases at least Q20
 Estimated insert size: 92352; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently consists of 52 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 6916: contig of 6916 bp in length
 6917 7016: gap of unknown length
 7017 11089: contig of 4073 bp in length
 11090 11189: gap of unknown length
 11190 14517: contig of 3328 bp in length
 14518 14617: gap of unknown length
 14618 17992: contig of 3375 bp in length
 17993 18092: gap of unknown length
 18093 21658: contig of 3566 bp in length
 21659 21758: gap of unknown length
 21759 25307: contig of 3549 bp in length
 25308 25407: gap of unknown length
 25408 27818: contig of 2411 bp in length
 27819 31526: contig of 3608 bp in length
 31527 31626: gap of unknown length
 31627 33967: contig of 2341 bp in length
 33968 34067: gap of unknown length
 34068 36642: contig of 2575 bp in length
 36643 36743: gap of unknown length
 36743 39205: contig of 2463 bp in length
 39206 39305: gap of unknown length
 39306 41281: contig of 1976 bp in length
 41282 41381: gap of unknown length
 41382 45065: contig of 3684 bp in length
 45066 45165: gap of unknown length
 45166 47143: contig of 1978 bp in length
 47144 47243: gap of unknown length
 47244 48499: contig of 1256 bp in length
 48500 50055: contig of 1456 bp in length
 50056 50155: gap of unknown length
 50156 53530: contig of 3375 bp in length
 53531 53630: gap of unknown length
 53631 55445: contig of 1815 bp in length
 55446 58015: contig of 2470 bp in length
 58016 58115: gap of unknown length
 58116 59712: contig of 1497 bp in length
 59713 61702: contig of 1990 bp in length
 61703 61802: gap of unknown length
 61803 63157: contig of 1355 bp in length
 63158 63257: gap of unknown length
 63258 64623: contig of 1366 bp in length
 64624 64723: gap of unknown length
 64724 65766: contig of 1043 bp in length
 65767 67869: gap of unknown length
 67869 67969: gap of 2003 bp in length
 67970 68560: gap of unknown length
 68560 69660: contig of 1591 bp in length
 69661 71980: contig of 2320 bp in length
 71981 72080: gap of unknown length

72081 73657: contig of 1577 bp in length
 73658 73757: gap of unknown length
 73758 75695: contig of 1938 bp in length
 75696 75795: gap of unknown length
 75796 78395: contig of 2600 bp in length
 78396 78495: gap of unknown length
 78496 79707: contig of 1212 bp in length
 79708 81374: contig of 1567 bp in length
 81375 81474: gap of unknown length
 81475 83163: contig of 1689 bp in length
 83164 83263: gap of unknown length
 83264 85474: contig of 2211 bp in length
 85475 85574: gap of unknown length
 85575 87264: contig of 1690 bp in length
 87265 87364: gap of unknown length
 87365 88370: contig of 1006 bp in length
 88371 88470: gap of unknown length
 88471 89719: contig of 1249 bp in length
 89720 89819: gap of unknown length
 89820 91531: contig of 1712 bp in length
 91532 91631: gap of unknown length
 91632 92853: contig of 1222 bp in length
 92854 92953: gap of unknown length
 92954 95061: contig of 2108 bp in length
 95062 95161: gap of unknown length
 95162 97151: contig of 1990 bp in length
 97152 97251: gap of unknown length
 97252 98627: contig of 1576 bp in length
 98628 98927: gap of unknown length
 98928 100300: contig of 1373 bp in length
 100301 100400: gap of unknown length
 100401 101467: contig of 1067 bp in length
 101468 101567: gap of unknown length
 101568 102772: contig of 1205 bp in length
 102773 102872: gap of unknown length
 102873 104128: contig of 1256 bp in length
 104129 104228: gap of unknown length
 104229 105514: contig of 1286 bp in length
 105515 105614: gap of unknown length
 105615 106797: contig of 1183 bp in length
 106798 106897: gap of unknown length
 106899 108212: contig of 1315 bp in length
 108213 108312: gap of unknown length
 108313 109627: contig of 1315 bp in length
 109628 109727: gap of unknown length
 109728 110921: contig of 1194 bp in length
 110922 111021: gap of unknown length
 111022 112730: contig of 1709 bp in length.

FEATURES
 source 1. 112730
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-5N18"
 BASE COUNT 30919 a 22306 c 22742 g 31565 t 5198 others
 ORIGIN

Query Match 36.7%; Score 339; DB 2; Length 112730;
 Best Local Similarity 62.6%; Pred. No. 2.2e-85;
 Matches 563; Conservative 0; Mismatches 330; Indels 6; Gaps 2;

QY 1 atgaatcaagaggttgtaactagttcatattatcttggcgtccacaaagctgaac 60
 DB 96341 ATGAATGGGACACTGGCTCTAGTTCATTCATTCGAGAAATGCTCACCTT 96282
 QY 61 cagggaattatctctctcttcttcattgctatctctgtgctttctcgcaaatg 120
 DB 96281 CGGG---TACCACTTCTCAGCTTCCTCCTCATATCATGCGCATTCACAGAAACCTG 96225
 QY 121 ctcatcatcatgccaataatctatagcaaaccttgcatagccatgatatgttttctt 180
 DB 96224 CTCATTATGTCAGTCAAGTCAAGTCCAGCCCTGCAATACCCCTATGACTTCTCTG 96165

[illegible]

OY 904 ttgcatttctgaacacta 923
|| | ||||| | | | |
Db 59336 TTCCTTTTCGAAATTA 59355

Search completed: June 21, 2002, 12:36:47
Job time: 8112 sec

RESULT	1
AA542323	
ID	AA542323 standard; cDNA; 924 BP.
XX	
AC	AA542323;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human cDNA encoding olfactory receptor AOLFRI24.
XX	
KW	Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant
KW	ss; food additive; cosmetic; fragrance; pharmaceutical additive.
XX	
OS	Homo sapiens.
XX	
PN	W0200168805-A2.
XX	
PD	20-SEP-2001.
XX	
PF	13-MAR-2001; 2001WO-US07771.
XX	
PR	13-MAR-2000; 2000US-0188914.
PR	24-MAR-2000; 2000US-0192033.
PR	12-APR-2000; 2000US-0198474.
PR	24-APR-2000; 2000US-0199335.
PR	26-MAY-2000; 2000US-0207702.
PR	23-JUN-2000; 2000US-0213849.
PR	16-AUG-2000; 2000US-0226534.
PR	07-SEP-2000; 2000US-0230732.
PR	07-FEB-2001; 2001US-0266662.
XX	
PA	(SENO-) SENOMYX INC.

XX Zozulya S;
 FI WPI: 2001-570867/64.
 DR P-PSDB: AAU24630.

XX Nucleic acids encoding human olfactory G protein-coupled receptors,
 PT useful for screening for compounds involved in olfactory sensation,
 PT where the compounds can be used in the food, pharmaceutical and
 PT cosmetic industries to customise odours -
 PS Claim 1; Page 135; 319pp; English.

XX The invention relates to nucleic acids encoding human olfactory
 CC receptors, OR, (a G protein-coupled receptor, GPCR). The ORs
 CC specifically recognise molecules, odorants, that elicit specific
 CC olfactory sensation. The human olfactory receptors and polynucleotides
 CC encoding them are useful for screening a library of chemical compounds
 CC for compounds that are involved in olfactory sensation. Modulators of
 CC their activity are useful for pharmacological and genetic modulation of
 CC olfactory signalling pathways. Therefore, they can be used in the food,
 CC pharmaceutical and cosmetic industries to customise odours and
 CC fragrances. The present sequence encodes a human olfactory receptor of
 CC the invention.

XX Sequence 924 BP; 219 A; 235 C; 184 G; 286 T; 0 other;

Query Match 99.7%; Score 920.8; DB 22; Length 924;
 Best Local Similarly 99.8%; Pred. No. 3.4e-276;
 Matches 922; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaatcacagcgtgtgaactgaatcatatctctggccctcccaaaaagcctgaacc 60
 DB 1 atgaatcacagcgtgtgaactgaatcatatctctggccctcccaaaaagcctgaacc 60
 OY 61 cagggaattatct 120
 DB 61 cagggaattatct 120
 OY 121 cctcatcatcatgtgcaaatatctatagcaaacctctgcatagccatgattgttctct 180
 DB 121 cctcatcatcatgtgcaaatatctatagcaaacctctgcatagccatgattgttctct 180
 OY 181 ctgacacatgctgtgtgtgacatcatctgcacaaagcatcatcaccgaagatgctgag 240
 DB 181 ctgacacatgctgtgtgtgacatcatctgcacaaagcatcatcaccgaagatgctgag 240
 OY 241 accatgctacatcagaataacattcatatgcaaggctgcatgtccagctctctctg 300
 DB 241 accatgctacatcagaataacattcatatgcaaggctgcatgtccagctctctctg 300
 OY 301 ttacatgtctctggagcgtgagatggtctcttcacacacatgscatagccgcat 360
 DB 301 ttacatgtctctggagcgtgagatggtctcttcacacacatgscatagccgcat 360
 OY 361 gtggcattgttctcctctctcatatcagtaactatgaaacacatggtgtagcc 420
 DB 361 gtggcattgttctcctctctcatatcagtaactatgaaacacatggtgtagcc 420
 OY 421 ttgtcagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 480
 DB 421 ttgtcagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 480
 OY 481 atgaggtgacttctctggtgccaacacacatgacacattctctctgtgagatacccca 540
 DB 481 atgaggtgacttctctggtgccaacacacatgacacattctctctgtgagatacccca 540
 OY 541 ttgtcagcttctcctctgtagccctgtaagaataagtgatgtagttagttagttagt 600
 DB 541 ttgtcagcttctcctctgtagccctgtaagaataagtgatgtagttagttagttagt 600
 OY 601 attaccctggcagcatagggaattattcttactgcatctctcctatgatttaccatgtt 660

DB 601 attaccctggcagcatagggaattattcttaccctgcatctcctatgatttaccatgtt 660
 OY 661 gctattctccgtatccagcagtagaaggaaggaagggcctctcaaatgctacatc 720
 DB 661 gctattctccgtatccagcagtagaaggaaggaagggcctctcaaatgctacatc 720
 OY 721 catctcagctgtgacacattacattctctctctctctctctctctctctctctctct 780
 DB 721 catctcagctgtgacacattacattctctctctctctctctctctctctctctctct 780
 OY 781 tccagctatcatcttgaagaagacaaggtgtgactgcaacatctactctgtgactcc 840
 DB 781 tccagctatcatcttgaagaagacaaggtgtgactgcaacatctactctgtgactcc 840
 OY 841 acattaaacccgattgtgacagcttccagaaataggaatgcaaggaattagaag 900
 DB 841 acattaaacccgattgtgacagcttccagaaataggaatgcaaggaattagaag 900
 OY 901 gtgttgattctctgaaacactag 924
 DB 901 gtgttgattctctgaaacactag 924

RESULT 2

AAH32037 standard; DNA: 921 BP.

AAH32037:

30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 610.

Human; olfactory receptor; OR; primary scent determination;
 secondary scent determination; polypeptide library; odour receptor;
 scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

WO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGI-) DIGISCENTS.

(YEDA) YEDA RES & DEV CO LTD.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI: 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 425; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides
 which encode polypeptides involved in olfactory sensation. The
 polynucleotides can be used in screening for olfactory agonists and
 antagonists. The methods allow for the determination of primary
 scents and the identification of the odour receptors used to detect
 these primary scents. The methods also enable determination of
 secondary scents and the identification of combinations of odour
 receptors that are involved in detecting such secondary scents.
 This enables the construction of a scent representation (also called
 a scent fingerprint or scent profile), which may be used to re-create
 and edit scents. Libraries of olfactory receptors are useful for
 determining the interaction pattern of a composition with the receptors.

CC and can be used for determining differences in the olfactory faculties
 CC of different individuals.

XX Sequence 921 BP, 218 A; 235 C; 183 G; 285 T; 0 other:

SO

Query Match 99.3%; Score 917.8; DB 22; Length 921;
 Best Local Similarity 99.8%; Pred. No. 2.9e-275;
 Matches 919; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 atgaatcacagcgtgtgaactgaagtcattatctcggcctcaccacaaagcgtgaactc 60
DB 1 atgaatcacagcgtgtgaactgaagtcattatctcggcctcaccacaaagcgtgaactc 60
QY 61 cagggaatattctctctctctctctctctctctctctctctctctctctctctctctct 120
DB 61 cagggaatattctctctctctctctctctctctctctctctctctctctctctctctct 120
QY 121 ctcatcatcatctgccaataatcctatagcaaacctctgcatagcccatgatttctcct 180
DB 121 ctcatcatcatctgccaataatcctatagcaaacctctgcatagcccatgatttctcct 180
QY 181 ctgacactgctgtgtgtgacatcctcgcacacaaagcatataccgaagatgctgag 240
DB 181 ctgacactgctgtgtgtgacatcctcgcacacaaagcatataccgaagatgctgag 240
QY 241 accatgctcaaatcagaataatcattcatatgtaggctgtagtccagctctctctg 300
DB 241 accatgctcaaatcagaataatcattcatatgtaggctgtagtccagctctctctg 300
QY 301 ttcaatgtctctcgtgagcgtgagatgttctctcaccacacatgacctatgaccgcat 360
DB 301 ttcaatgtctctcgtgagcgtgagatgttctctcaccacacatgacctatgaccgcat 360
QY 361 gtgacattgttctccctcctcattacagtaactgtatgaacacacatattgtgtagcc 420
DB 361 gtgacattgttctccctcctcattacagtaactgtatgaacacacatattgtgtagcc 420
QY 421 ttgctcagcatgtagtcatgtagtcatgtagtcatgtagtcatgtagtcatgtagtcatg 480
DB 421 ttgctcagcatgtagtcatgtagtcatgtagtcatgtagtcatgtagtcatgtagtcatg 480
QY 481 atgaggttgaacttctcgtgagcgaacacacatgtagtcatgtagtcatgtagtcatg 540
DB 481 atgaggttgaacttctcgtgagcgaacacacatgtagtcatgtagtcatgtagtcatg 540
QY 541 ttgctgcttctgctcgtgagcgtgagatgaatgaatgagtgtagtctgtagt 600
DB 541 ttgctgcttctgctcgtgagcgtgagatgaatgaatgagtgtagtctgtagt 600
QY 601 attaccctgagcagaatgagtagtcttctcctcctcctcctcctcctcctcctcctcct 660
DB 601 attaccctgagcagaatgagtagtcttctcctcctcctcctcctcctcctcctcctcct 660
QY 661 gcatattcctcgtatccgacagtagaagcaaggaaggaagccttcaaacatgctcatct 720
DB 661 gcatattcctcgtatccgacagtagaagcaaggaaggaagccttcaaacatgctcatct 720
QY 721 catctcagatgtagccttactatctcctcgttaacttcaacatataccgacctgct 780
DB 721 catctcagatgtagccttactatctcctcgttaacttcaacatataccgacctgct 780
QY 781 tccagctatactttgaaagagacaggtgtgtagtgcattacactctgtgtagctcc 840
DB 781 tccagctatactttgaaagagacaggtgtgtagtgcattacactctgtgtagctcc 840
QY 841 acattaaaccgagtgtagcagcttccagaaatagagatgcaagcaggaattaggaag 900
DB 841 acattaaaccgagtgtagcagcttccagaaatagagatgcaagcaggaattaggaag 900
QY 901 gtgttgcatttctgaacac 921
DB 901 gtgttgcatttctgaacac 921
  
```

RESULT 3

AS17179 standard; cDNA: 798 BP.

XX

AS17179;

14-FEB-2002 (first entry)

DT

XX

DE

XX

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

Human cDNA encoding a novel G protein-coupled receptor, NOV6.
 Human: ss; G protein-coupled receptor; GPCR; NOV6; cardiant;
 antidiabetic; cytotonic; analgesic; antisthmatic; nootropic;
 cardiomyopathy; atherosclerosis; infection; pain; anorexia; bulimia;
 asthma; neurological disorder; Parkinson's disease; stroke;
 Alzheimer's disease; multiple sclerosis; leisional psoriatic skin;
 ischaemia; cirrhotic hepatitis; acute pancreatitis; diabetes; cancer;
 angiogenesis; obesity; olfactory disorder; chromosome 11.
 Homo sapiens.
 Key Location/Qualifiers
 5'UTR 1..23
 CDS 24..782
 3'UTR 783..798
 WO200177177-A2.
 18-OCT-2001.
 11-APR-2001; 2001WO-US11901.
 11-APR-2000; 2000US-19594P.
 11-APR-2000; 2000US-19653P.
 26-APR-2000; 2000US-19902P.
 26-APR-2000; 2000US-19964P.
 27-APR-2000; 2000US-19955P.
 27-APR-2000; 2000US-19956P.
 27-APR-2000; 2000US-200176P.
 17-JUL-2000; 2000US-21895P.
 25-JUL-2000; 2000US-22064P.
 04-JAN-2001; 2001US-259641P.
 29-JAN-2001; 2001US-264851P.
 14-FEB-2001; 2001US-268567P.
 13-MAR-2001; 2001US-0220644.
 (CURA-) CURAGEN CORP.
 Shenoy S, Gangoli EAG, Rastelli L, Smithson G, Padigaru M;
 Vernet CM, Wolenc AR, Casman SO, Tchernev VT, Szekeres ES;
 Gorse W, Alsobrook JP, Burgess CE;
 WPI: 2002-041291/05.
 P-PSDB; AMU11099.
 Human G-protein coupled receptors, NOV1-12, useful for diagnosis and
 treatment of e.g. cardiomyopathy, also in screening for specific
 modulators -
 Claim 11; Page 30; 171pp; English.
 The invention relates to Human G-protein coupled receptor (GPCR)
 polypeptides, designated NOV1-12, nucleic acids encoding them,
 variants of them, a vector containing the nucleic acid, a host cell
 containing the vector, a method for identifying modulators of the GPCRs

OY	4	aatacagcgcttgtaactgagatlcattatctctgagctccacaaagaagcttgaaatccgag	63
Db	321	aaccagaaagcttgtaacccagatcctctgaaagagcttcttgagaccacgaataacgg	380
OY	64	ggaattactctcccttttctctaatgtgtactgtgtgtgtgtcttcccggaacatgtctc	123
Db	381	gtgtcttaattaaagctgttc	440
OY	124	atcattcatgtccaaatctatagacaacactgtcaacagcccatgatatgtttctctctg	183
Db	441	atccactctggccatcaagcttcaacacctgtggtccacagctctctatgtactttcttacc	500
OY	184	acacgtgctgtgttggaattactctgcacaaagaacacacatcaacgaagaatgtcggagcc	243
Db	501	aactctgtgcttaactatggaactatctctgcaacctcttcacatcagcccaagcgctgacagt	560
OY	244	atgtctaacatcagaanaatcaattctcatatgatcaggtctgatatgtccagctctctgttc	303
Db	561	ctggtctcggaagaaagagcccactctctcaggggctgtgcatatgagccagctctattcttc	620
OY	304	acatgtctctcgtggagctgagatgtgtctctctcaccacacatgtgcctatgaacgctatgtg	363
Db	621	acgtgggcgcgaatccccaagatcgctgcgtccctcaacggtatcgtctatgaccgtataga	680
OY	364	ggcaattgtttccctctctcaattacagtaactatatagaacacacatgtgtgttaaccttg	423
Db	681	ggcaatctgcacacccgcgtgatctacaaagaaatgataatgagaagaagtgtctgcagcggtcg	740
OY	424	ctgaagatgtgcatgtgctatgtgacgtacacaaatctctgtgtgcacaacagctcttacaatg	483
Db	741	ggccacagcgtgtgtgctctgctctgctgcgcgttcaacaagccatccacacgggctgtgaatgtg	800
OY	484	aggttgacttctctgtgvggccaacaacatgtgaccactctctctgtagatacccccatgt	543
Db	801	cgcttgatattctgtgtgccccaaagtcatattccattctctctctctctctctctctctg	860
OY	544	ctggtctgtgtcctgtagccctgtaagaatcaatgaagatgtgatagtgtgtctgataatt	603
Db	861	ctgctctctctctctcagctccacactaagctcaacggtgtcatagattgtctctcggagatgt	920
OY	604	aaccttgccaatgagggacttattctctacactgcatactccatagttttatcatatgtgtct	663
Db	921	ttctacggcatatgaaactctctctatgacccctcgcgtccatgtctctcatcgcgtctcagc	980
OY	664	attctccgtatccgcagatgaaagagcagaaggaagcgctctcacaatgtgtatctcat	723
Db	981	atctcgaaagtggaagacgtcctgtgggggagagaagcgtctctccactcgtctcttccac	1040
OY	724	ctcacagtggtgaccccttactatctctctctctctctctctctctctctctctctctctc	783
Db	1041	ctcacagtggtggtgcatgtatataacacgggtgtctctctacgcctacataaagccgctctct	1100
OY	784	agctataacttggaaagacagaatgtgtgtagctgacatactactctgtgtactccaca	843
Db	1101	ggctacagcgacgggaaggaagcaggtgtgctggcgtctgtatcaactatgtgtgagtctacc	1160
OY	844	ttaaaccgaatgtgttacagctctccaagatagggagatgycggagcaggaattaggaaagt	903
Db	1161	ctcaaacccctcatctatacttlttgaaacaaagagaggtcaaaagcagccctcagaagatt	1220
OY	904	tttgactttctgnaacacta	923
Db	1221	ttctcccttctctcagaattat	1240

XX	30-JUL-2001	(first entry)
XX		
XX	AAH31706;	
XX		
XX	AAH31706	standard; DNA; 924 BP
AAH31706		
RESULT	6	

DE	Human olfactory receptor polynucleotide, SEQ ID NO: 279.
XX	Human; olfactory receptor; OR; primary scent determination;
KW	secondary scent determination; polypeptide library; odour receptor;
KW	scent profile; scent fingerprint; scent representation; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200127158-A2.
XX	
PD	19-APR-2001.
XX	
PF	06-OCT-2000; 2000WO-US27582.
XX	
PR	08-OCT-1999; 99US-0158615.
XX	
PR	24-FEB-2000; 2000US-0184809.
XX	
PA	(DIGI-) YEDA RES & DEV CO LTD.
XX	
PI	Belenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX	
DR	WPI: 2001-290713/30.
XX	
PT	New polynucleotides which encode polypeptides involved in olfactory
XX	sensation for identifying olfactory agonists and antagonists -
XX	Claim 8; Page 289; 1857pp; English.
PS	
XX	
CC	The present sequence is one of a number of isolated polynucleotides
XX	which encode polypeptides involved in olfactory sensation. The
CC	polynucleotides can be used in screening for olfactory agonists and
XX	antagonists. The methods allow for the determination of primary
CC	scents and the identification of the odour receptors used to detect
XX	these primary scents. The methods also enable determination of
CC	secondary scents and the identification of combinations of odour
XX	receptors that are involved in detecting such secondary scents.
CC	This enables the construction of a scent representation (also called
XX	a scent fingerprint or scent profile), which may be used to re-create
CC	and edit scents. Libraries of olfactory receptors are useful for
XX	determining the interaction pattern of a composition with the receptors,
CC	and can be used for determining differences in the olfactory faculties
XX	of different individuals.
CC	
XX	Sequence 924 BP; 170 A; 293 C; 219 G; 242 T; 0 other;
XX	
Query Match	33.1%; Score 305.4; DB 22; Length 924;
Best Local Similarity	58.4%; Pred. No. 1,5e-84;
Matches	534; Conservative 0; Mismatches 381; Indels 0; Gaps 0.
OY	4 atcaacagcgtttgaactgagttatattcttggtcctacccaagaacgtgaactcag 63
DB	10 aaccagcagtttggtaaacgagttcaactcgtcagggcttttcggagccacgaatacgg 69
OY	64 ggaatacttcctcttttttctcaatgtatcttvtggtctttctcggacaacatgc 123
DB	70 ggttcttattatagtttcttcctcttcaactcgttgggacctacaggaatgctc 129
OY	124 atcatcatgccaataatctatagaacaccccttgatagcgccatgtatgttttccctc 183
DB	130 atcaacttgcacatcagttcaaccttgggtcgaagctcactatgtatctttctactc 189
OY	184 acatggtcgtgttggtgacatcatctgcacaacaagcatcataccgaatgtcggacc 243
DB	190 aacttgcactatggaactatctcgaacctcttcacatcgaaccaagcgttcgacgt 249
OY	244 atgtcaacatcgaataatccattcatatgcaggctgcagtccacgtctttcttctc 303
DB	250 ctgtgttcggaagaggtcctactcctacagggtgtcgtatgcacgacttactcctc 309
OY	304 acatgtctctcggagcgtgagatgttctctcacacacatgagcctatgacgtatgt 363

DR	WPI: 2001-290713/30.
XX	
PT	New polynucleotides which encode polypeptides involved in olfactory
XX	sensation for identifying olfactory agonists and antagonists -
PS	
XX	Claim 8: Page 424; 1857pp; English.
CC	The present sequence is one of a number of isolated polynucleotides
CC	which encode polypeptides involved in olfactory sensation. The
CC	polynucleotides can be used in screening for olfactory agonists and
CC	antagonists. The methods allow for the determination of primary
CC	scents and the identification of the odour receptors used to detect
CC	these primary scents. The methods also enable determination of
CC	secondary scents and the identification of combinations of odour
CC	receptors that are involved in detecting such secondary scents.
CC	This enables the construction of a scent representation (also called
CC	a scent fingerprint or scent profile), which may be used to re-create
CC	and edit scents. Libraries of olfactory receptors are useful for
CC	determining the interaction pattern of a composition with the receptors,
CC	and can be used for determining differences in the olfactory faculties
CC	of different individuals.
XX	
SQ	Sequence 972 BP; 206 A; 252 C; 210 G; 304 T; 0 other;
	Query Match 28.9%; Score 267.4; DB 22; Length 972;
	Best Local Similarity 56.0%; Pred. No. 1.1e-72;
	Matches 505; Conservative 0; Mismatches 396; Indels 0; Gaps 0;
QY	1 atgaatcacaggtgtgtactgagtcatacttacccttcggccttaccacaaagactgaatc 60
DB	13 aggaacaacacaaatgtgacaanaattccttcctcgttggaacttcagaccatccaaatg 72
QY	61 cagggaatatctctcctctttttctcaatgtctatctcttggccttcttcggcaaatg 120
DB	73 aagatttctcttctaagtgatttcttggggcctacccttcgaagtggtgcgggaactta 132
QY	121 ctccatcatatgccaanaatctatagacaacacttgacatagcccatgtatgttctct 180
DB	133 agcccatctgcccattataagatgactctcaaccgcgaatgccaatgtactctctc 192
QY	181 ctgacatcgtgctgtgttggaacatctgcacaaagaacatatacgaagaatgctggg 240
DB	193 agtaaccgttccctctctgacatctgcgtatgtgtctccacacgcccctaagaatgtcgt 252
QY	241 accatgccaacacgaagaataacatcttcatatgagcgtgatagtgccacgtctctg 300
DB	253 gacatcatcacaagagcaaaaacacttcccttctgttgcgtgcgcacccagaactctgc 312
QY	301 ttcaatatgctccttgaggcgttgagatgtgtctcttcacacacatgacctatgaccgtat 360
DB	313 ttctgtggagatgggctgactgaatgtcttctcctgcgcagtatgctatgaccggtat 372
QY	361 gtggccattgtttccctctcatatcagtgactatatagaaccaccataatgtgtgagcc 420
DB	373 gctgaatctgaaaccccttgcttaacacagtcctcatatccacaacacattgtttaag 432
QY	421 ttgctccagcatgtcatgtgcatatgcagtcacaaactctctggtgcacaagctctta 480
DB	433 atggtggttgaggcctatggtggtgagtatctcttagtcttctcatatgaaacatactgtc 492
QY	481 atgagatgtgaacttctgtgggccaacaacacatgacaactctctctgtgagatacccca 540
DB	493 tatcagcatgattttcttggggcccatatgatacaacacttctctgtgacctctcccca 552
QY	541 ttgctgctctgtctctgtagcccttgtaagaatcaatggatggatgtagtattgtcgtat 600
DB	553 gtcccggtcgtgctcgtctgataccttcaacacgaggtggatgaaccttcaatgaagtcag 612
QY	601 ataccctggcatatggggaacttatacttcaactgacatccatcctatggtttatcatatgt 660
DB	613 gtgtcgttggataatagtgctctgtgtagtggtcttccatctctatagtgtaataatgtgtct 672

CC	for compounds that
CC	for compounds that

5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528

Db 913 g 913

RESULT 9
AADI9150 standard; cDNA: 1050 BP.

AC AADI9150:
18-DEC-2001 (first entry)

Human G-protein coupled receptor 17 (GPCR17) cDNA.

Human: G-protein coupled receptor 17; GPCR17; cardiomyopathy; vaccine; atherosclerosis; diabetes; cardiomyopathy; cancer; obesity; pain; diabetes mellitus; anorexia; cachexia; cardiomyopathy; atherosclerosis; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; immune disorder; haematopoietic disorders; retinal disorder; HIV; human immunodeficiency virus; adenocarcinoma; bulimia; asthma; ulcer; angina pectoris; hypertension; hyperextension; Crohn's disease; anxiety; multiple sclerosis; schizophrenia; dementia; mental retardation; gene therapy; osteoporosis; urinary retention; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FH 5'UTR 1..30
FT CDS /*tag= a
FT 31..1005
FT /*tag= b
FT /product= "Human GPCR17 protein"
FT sig_peptide /*tag= c
FT 31..153
FT mat_peptide /*tag= d
FT 154..1002
FT /product= "Mature GPCR17 protein"
FT 1006..1050
FT 3'UTR /*tag= e

XX WO200181378-A2.
XX 01-NOV-2001.
XX 27-APR-2001; 2001WO-US13680.
XX 27-APR-2000; 2000US-199947P.
XX 27-APR-2000; 2000US-199947P.
XX 14-AUG-2000; 2000US-275226P.
XX 18-DEC-2000; 2000US-256524P.
XX 18-DEC-2000; 2000US-256524P.
XX 22-DEC-2000; 2000US-258159P.
XX 28-DEC-2000; 2000US-258159P.
XX 28-DEC-2000; 2000US-258159P.
XX 04-JAN-2001; 2001US-259659P.
XX 13-MAR-2001; 2001US-0275226.
XX (CURA-) CURAGEN CORP.
XX Padigaru M, Mishra V, Spytek KA, Grosse WM, Szekeres ES;
XX Alsobrook JP, Burgess CE, Casman SD, Lepley DM, Gangoli EA;
XX MacDougall JR, Smithson G;
XX WPI; 2001-611739/70.
XX P-PSDB; AAE11913.
XX G-protein coupled receptor polypeptides and NAs useful for
XX preventing, diagnosing and treating cardiomyopathy, atherosclerosis,
XX cancers and diabetes -
XX Claim 5; Page 64; 242pp; English.
XX The present sequence is human G-protein coupled receptor-17 (GPCR-17)

CC cDNA. GPCR protein and DNA may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate GPCR expression,
CC obesity, diabetes mellitus, anorexia, cachexia, cardiomyopathy, pain,
CC atherosclerosis, neurodegenerative disorders (Alzheimer's disease,
CC Parkinson's disease, Huntington's disease), bulimia, immune disorder,
CC haematopoietic disorders, disorders related to cell signal processing
CC and metabolic pathway modulation, retinal disorder (photoreception),
CC bacterial, fungal, protozoal and viral infections (HIV); cancer (neoplasm
CC adenocarcinoma); angina pectoris, hypertension, hyperextension, asthma,
CC Crohn's disease, multiple sclerosis, ulcers, neurological disorders
CC (dementia, mental retardation, schizophrenia, anxiety); acute heart
CC failure, osteoporosis, myocardial infarction and urinary retention.

Sequence 1050 BP; 233 A; 262 C; 229 G; 326 T; 0 other:

Query Match 28.9%; Score 267.4; DB 22; Length 1050;
Best Local Similarity 56.0%; Pred. No. 1.1e-72;
Matches 505; Conservative 0; Mismatches 396; Indels 0; Gaps 0;

QY 1 atgaatcacagcgttgtaactgattcatattctctgggctcaccacaagcctggaactc 60
Db 43 aggaacaacacaaattggaacaaattcattccctcgggacttcagacacccccaatg 102

QY 61 cagggaattatctctctctcttttctcattgtctatctgttgcttcttcggacaatg 120
Db 103 aagattctcttcttcatgtattctcgggctcaccctcctgagcttgccggaaacta 162

QY 121 ctcatcatcattgccaataatcctatagcaacacctgacatgacgccaatgtattctctt 180
Db 163 agcctcatgcccctcattagaatgacctcaccgccaatgccaatgtactctctc 222

QY 181 ctgacacgtgctgtgtgacatctctgcacaagaagacatacagaagatgctggg 240
Db 223 agtaacctgctctctctcgtgacatctgctgtgtcctccaccggccccaagatgct 282

QY 241 accatgctaacacacagaataatccattcatatagagcgtgacgtgccagctctctg 300
Db 283 gacatcatcacagagagaagaacacattctcttctgttggtggccaactgaacttctg 342

QY 301 ttacatgtctctggtgagatgagtgtctctcaccacacatgagccatgacggcat 360
Db 343 tctctggaatggtggtgacgtgaatgacttctcctcggcagcatggtccatgaaccgcat 402

QY 361 gtgcacattgttccctctcattacagtaactatattgaaccacatatgtgttagcc 420
Db 403 gctgcaattctgaacacctctgcttacaacagccctcatctccaacattgtttaag 462

QY 421 ttgctcagcatggtcatgtgctatgcaatccctcctggtgacacagctctatc 480
Db 463 atgtgtgttggtggtcctatgtggtatgcttcttcttcttcttcttcttcttcttct 522

QY 481 atgaagttgaccttctgtgggccaacacctatgacaccttcttctgtgagatacccca 540
Db 523 tatcagcatgattctctgtgggcccctatgatacaacaccttcttctgtgacctctcca 582

QY 541 ttgctgctgttctgtctgagccctgtaagaatcaatgagatggtgtatgtctgat 600
Db 583 gtccgtgctctgtctgtctgtctatccttcaacacagcggtgtgtaacctctatagct 642

QY 601 attaccctggccaataggggacttattcttaccatgctctcctatgatttcatgct 660
Db 643 gtctgtgtggaatgctgtgtgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 702

QY 661 gtattctccgatactcgcacagatgaagaaggaagagccttctcaaatgctcatc 720
Db 703 gctgtgtggaatgactagctcagctacagtaggaagaagagccttcaagactgtgctct 762

QY 721 catccacagtgtaaccttactatctctcgtgaactcaacatcatccgctgct 780
Db 763 cactgactgctgtgacctctctctatgttctgtgattcttcatgtataatgacgacct 822

QY 781 tccagcatataatttgaaagacaaagtggtagctgacactatatactctgtgactcc 840

Db	823	tcgaagctactccctaacaacgagcgaagtggtgtccatattttagcttggtgaccc	882
Oy	841	acataaacccgattggtgacagctccagataggagatgcaagccaggaattaggaag	900
Db	883	gtgtggaatcccatcatcatcacaatttttagagataagagattaaatgycatgagaa	942
Oy	901	g 901	
Db	943	g 943	
RESULT	10		
ID	AAH31995		
AC	AAH31995	standard; DNA; 930 BP.	
XX	AAH31995;		
XX	30-JUL-2001	(first entry)	
DE	Human olfactory receptor polynucleotide, SEQ ID NO: 568.		
XX	Human: olfactory receptor; OR; primary scent determination;		
KW	secondary scent determination; polypeptide library; odour receptor;		
KW	scent profile; scent fingerprint; scent representation; ds.		
XX	Homo sapiens.		
OS	MO200127158-A2.		
PN	19-APR-2001.		
PD	06-OCT-2000; 2000MO-US27582.		
PE	08-OCT-1999; 99US-0158615.		
PR	24-FEB-2000; 2000US-0184809.		
XX	(DIGI-) DIGISCENTS.		
PA	(YEDA) YEDA RES & DEV CO LTD.		
PI	Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;		
DR	WPI: 2001-290713/30.		
PT	New polynucleotides which encode polypeptides involved in olfactory		
PS	sensation for identifying olfactory agonists and antagonists -		
XX	Claim 8; Page 407; 1857Pp; English.		
XX	The present sequence is one of a number of isolated polynucleotides		
CC	which encode polypeptides involved in olfactory sensation. The		
CC	polynucleotides can be used in screening for olfactory agonists and		
CC	antagonists. The methods allow for the determination of primary		
CC	scent and the identification of the odour receptors used to detect		
CC	these primary scents. The methods also enable determination of		
CC	secondary scents and the identification of combinations of odour		
CC	receptors that are involved in detecting such secondary scents.		
CC	This enables the construction of a scent representation (also called		
CC	a scent fingerprint or scent profile), which may be used to re-create		
CC	and edit scents. Libraries of olfactory receptors are useful for		
CC	determining the interaction pattern of a composition with the receptors		
CC	and can be used for determining differences in the olfactory faculties		
CC	of different individuals.		
XX	Sequence 930 BP; 204 A; 259 C; 178 G; 289 T; 0 other;		

Query Match	28.23;	Score 260.8;	DB 22;	Length 930;
Best Local Similarity	55.78;	Pred. No. 1.2e-70;		
Matches 499;	Conservative	0;	Mismatches 397;	Indels 0;
				Gaps 0;
14 ttgttaactgaattcattatcttctgggcctccacccaaaagcctgaaatccaggaattatctc	73			

Db	23	tggtagctaggttattcctcatcgtatcggtacatcccaaggaccctccagatgtagtactctct	82
OY	74	tccctctttttctcatctgtctatctctgtgaccttctccggacaacatgctcatcatctg	133
Db	83	tcgtgtgcttccctcatagtttaacctggttaatgtagtggggaataatltggtatgtatcc	142
OY	134	ccaaatctctatgacaacacctgmatagcccatatgatgtttctctctgcaactgctg	193
Db	143	tgtattacaacagacactcagcttcacacacccatgtaatttttctctcgaacctctcc	202
OY	194	ttgtgacacatctgcacacacaagaatatalacgaaagtctggtggacatgtcaat	253
Db	203	ttgtgtaccctgggtactactctctcaagcatctggcccccaagatgctggtctgacttccaa	262
OY	254	cagaanaatcacattcatatgatcagctgtcatgtccagctctctctgttcaaatgtctc	313
Db	263	atccacaagtattatctctctctccagctgtgccaacccagttgtctttttagtatttg	322
OY	314	tgggaagttagatgagttctctcttcacacacatggtccctatgacccgtatgtggtccatt	373
Db	323	tgtgattcctgagttgtaatgtctctgcgcagccatggtccctatgtctgttttggtccatt	382
OY	374	tccctcttcatatcagtaactattatgaaacacacatatgtgtgtagctgtgtccagacgg	433
Db	383	gaccctccacaatatagacacctcatgtcccaagaggtctgtctgtgtctcattcagctggct	442
OY	434	tcatgtgctattgcagtgtaaccaattccctgggtgacacacagctcttattcagaaagttagt	493
Db	443	cttaacctgtgcgtctcgtgtgtagtttagttagccacaactaaccctcaccttcagccagtt	502
OY	494	tctgttggcccaaacacacattgacacctctctctgtgtagatataccccaattgctgtgctt	553
Db	503	actgtgttccaatatacatcatcaatcttctctgcgaaacccacacacatctgtgacctct	562
OY	554	ccgtttagccctgtaagaatcataggttgaatgtgtatgtgtgtgtgtatataccctggcca	613
Db	563	ctgtctcagacacactacatcagtgtagatctgtctcttcaagtctgtgtgtctcaattgaat	622
OY	614	tagggaacttattattcttaacctgcacatctctcctaaggtttatcatatgtgtctattctcgra	673
Db	623	tcagacacatcctcatctcatctctcctcatatcccttattccttgttgaatcataga	682
OY	674	tccgacacgttgaaggacgaaggagggcccttccaatgatctcatatctcatctacagtagt	733
Db	683	tgcgttcgtcgaaagccgccttaaggcttctctcacactgtcggtgtctaaccttactgga	742
OY	734	tgaccttactatctctcgttaatactaacactatatacgcgcctgtctcaggtataact	793
Db	743	tcacctctctatgacagctcagctcagttatgtatccggaaggccaacaaatcaagactccc	802
OY	794	ttgaaagagacaaagtgtgtagctgcactctatactctgtgactccacaataaaccca	853
Db	803	tggaccacgaacaagtgggctcctgtgtctctacaaggtatcatcccatgtataatccct	862
OY	854	tgtgtgacgcttccagaatagtggaatgacggcaggaattaggaaggtgtttgca	909
Db	863	tgtatctacagtttggtggacaagaagtgtgaaagctgtcttcaaaaagctaatltgga	918
RESULT 11			
AAAS1177			
ID AAAS1177 standard; cDNA; 669 BP.			
XX AAAS1177;			
XX AC			
XX DT 26-SEP-2000 (first entry)			
XX DE			
XX XX Murine olfactory receptor ligand-binding region cDNA 3.			
XX KW Degenerate primer; ligand-binding region; olfactory receptor;			
XX human rhodopsin receptor; odorant; toxicity; olfactory response; ss.			
XX			

OS	Mus musculus.
XX	
FH	Key Location/Qualifiers
EH	1..669
FT	CDS
FT	/tag= A
FT	/product= Olfactory_receptor_ligand_binding_protein
FT	/partial
XX	
XX	WO200035274-A1.
XX	
PD	22-JUN-2000.
XX	
PF	17-DEC-1999; 99WO-US30221.
XX	
PR	17-DEC-1998; 98US-0112605.
XX	
PA	(UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX	
PI	Reed RR, Krautwurst D, Yau KW;
XX	
DR	WPI: 2000-431471/37.
XX	P-PsDB: AA96664.
XX	
PT	Primer pair for amplifying olfactory receptor nucleic acid, useful e.g.
XX	for producing receptor libraries used in e.g. screening odorants for
XX	toxicity
PS	Claim 18; Page 31; 61pp; English.
XX	
CC	AA51175-93 encode murine odorant/ligand binding regions of olfactory
CC	receptors. They were amplified using degenerate primers shown in
CC	AA51163-64.
CC	The primers can be used to generate a library of OR-LBR comprising the
CC	nucleic acids encoding a 7-TMD protein comprise a sequence encoding an
CC	N-terminal plasma membrane translocation domain (especially the sequence
CC	shown in AA96658), a first TMD and the OR-LBR. The translocation domain
CC	was initially derived from the N-terminus of the human rhodopsin
CC	receptor. LBR produced by amplification with the primers are useful for
CC	generating new odorants, to screen for toxicity or therapeutic activity
CC	in odorants, and altering an animal's olfactory response (claimed).
XX	
XX	
SQ	Sequence 669 BP; 131 A; 200 C; 142 G; 196 T; 0 other:
	Query Match 28.1%; Score 259.8; DB 21; Length 669;
	Best Local Similarity 62.0%; Pred. NO. 2e-70;
	Matches 411; Conservative 0; Mismatches 252; Indels 0; Gaps 0;
OY	187 ctgagctgttgatggacatcatcgcacaagaacatcacgaagtgtgaggacatg 246
Ob	7 ctggccaccactgatattatctcgacccttcctctgctgcgcccaagcgctgttgtcta 66
OY	247 ctaaacatacgaanaaaataacatttcaatatgcaggtgcattgtcccagctctttcttta 306
Dd	67 ctatctggagaaaacacacatctcctttaaagggtgcatgagccaagctcttcttcctgtg 126
OY	307 tggctctcggagagcgagatggtctcttcacaccacatgagccttatgacgcctatgtggc 366
Dd	127 tggctcctgctcttcagagctgctgctgtcgtctcaagtcattgagcctaatgtggcc 186
OY	367 attgttccctctcatlatacaglaactatltgaaccaacacatagtgtgtagaccttgctc 426
Dd	187 atctgcttccccctgactacacagctctagaattgagccacagttgttggggctctggcc 246
OY	427 agcattgcatgagcgcattgcgcagtcaccaattcctgtgggtgcaacagcctctatcatgag 486
Dd	247 atgggtgtatggttcatctgtgctctgtaagatcatlatacacaacggtctgtatgacacgg 306
OY	487 ttgacttctcttgggccaacacatltgacaccttctctgtgagatatcccccatgtgctg 546
Dd	307 ctgtcatctctgtgagcccaaggtcatcaccacatctctctctgtgtgagatcccccatctt 366

QY	547	gcttgcgtccctgtagccctgtaagaatcaatgaagtgatgagtgatgtagtgcgcgaattaac	600
Db	367	ctgcctccctgtagcccaatacgtataacacagcatatgactctaaatagaacagtcttc	428
QY	607	ctgcgcataagaggaacttatctctacccgcatactccctatggtttatcaatggttcatt	666
Db	427	tatggaagcatcaatattgtgtcttactctactatccctatgctgcattatgccaagcatc	488
QY	667	ctcgcgtatccgcacagtagaagaagcaagaagagcctctcaaatgactcatcattc	728
Db	487	ctgcgcacatgcgtctcgtcgcgaaggaagagcctttctactccgtctcatccacctc	548
QY	727	aacagtgtagacccttactatctctctgtaattctaacactatccgcctcgtccagc	788
Db	547	atcgttgctctctgtagtactaccatctctgttctgtgcctatgtaacgccctgcacagc	606
QY	787	tatacatcttgaagaagacaaggtgtgtagctgcactctatactctgtgacctccacata	848
Db	607	tatagcccaagaagaagaagaagtataccctctgtgtgtactatctccatcagcccaacctg	666
QY	847	aac	849
Db	667	aac	669
RESULT 12			
AAH31926			
ID	AAH31926 standard; DNA; 975 BP.		
XX			
AC	AAH31926;		
XX			
DT	30-JUL-2001 (first entry)		
XX			
DE	Human olfactory receptor polynucleotide, SEQ ID NO: 499.		
XX			
KM	Human: olfactory receptor; OR: primary scent determination;		
KM	secondary scent determination; polypeptide library; odour receptor;		
KM	scent profile; scent fingerprint; scent representation; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200127158-A2.		
XX			
PD	19-APR-2001.		
XX			
PF	06-OCT-2000; 2000WO-US27582.		
XX			
PR	08-OCT-1999; 99US-0158615.		
PR	24-FEB-2000; 2000US-0184809.		
XX			
PA	(DIGI-) DIGISCENTS.		
PA	(YEDA) YEDA RES & DEV CO LTD.		
XX			
PI	Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;		
XX			
DR	WPI; 2001-290713/30.		
XX			
PT	New polynucleotides which encode polypeptides involved in olfactory		
PT	sensation for identifying olfactory agonists and antagonists -		
XX			
PS	Claim 8; Page 380; 1857pp; English.		
XX			
CC	The present sequence is one of a number of isolated polynucleotides		
CC	which encode polypeptides involved in olfactory sensation. The		
CC	polynucleotides can be used in screening for olfactory agonists and		
CC	antagonists. The methods allow for the determination of primary		
CC	scents and the identification of the odour receptors used to detect		
CC	these primary scents. The methods also enable determination of		
CC	secondary scents and the identification of combinations of odour		
CC	receptors that are involved in detecting such secondary scents.		
CC	This enables the construction of a scent representation (also called		
CC	a scent fingerprint or scent profile), which may be used to re-create		
CC	and edit scents. Libraries of olfactory receptors are useful for		

Db 172 atcattgtgctacagcttgatgaactcattacaccacatgatactctcttcgcc 231
 QY 184 acactgtgctgttggaactatctgcacaaagacataaccaggaatgctgggacc 243
 Db 232 aatctactcttgctgatacttccatcttccaaactcagtcgccaaatgctggaat 291
 QY 244 atgctacacacgaataatcattcattatgacagctgagctgacagctctctgttc 303
 Db 292 attcaaaccaagaatcaatccatctctatagagctgacacaaagatgatacttct 351
 QY 304 acatgtctctcggagctgagatggtctctccacacacatgagctatgacagctatg 363
 Db 352 atgtgttctgctgcatcacaattgctcttgaggacatgagctatgacacattgctg 411
 QY 364 gccattgttccctcttattacagatactatgataaacacacatggtgtgagctgtg 423
 Db 412 gcatctgacaccccttgatatacattcctcagcggccaggttgacattgtctc 471
 QY 424 ctcaagatgctatgctatgcaagtcacacattcctggtgacacagctctatcatg 483
 Db 472 acagctatctcattgtctctcagtaataatattgtctctgacacaccccttgcacat 531
 QY 484 aggttgactctctgtggccaacacacattgacacactctctctgtgagataccacatg 543
 Db 532 caattgctctctgttaacacacacactctccacacactctctctgtaacttgccctctg 591
 QY 544 ctggcttctctgtgacccctgtaagaatgagtgatggtgtatgttctgataat 603
 Db 592 ctcaaacctgctctgtcagatacatgacatgacatgactgtgttctgttctgtggtta 651
 QY 604 accctggacataggaggaactattcttaccctcactcactgattatgattctgttgc 663
 Db 652 tcaattatcattctcccttccactcagctctcttctctatgctgacatcagaact 711
 QY 664 attctcgtatccgcacagtagaaggaaggaagggcctctccacatgctatcatcat 723
 Db 712 gctctgagatgattcttccacacaggaagtagaagggcctctccacatgctgctcacc 771
 QY 724 ctcaacagtggtgaccccttactatctctctgttaacttaactatactccgcccctgtcc 783
 Db 772 ctgaacagtgatctactgtcttccaggaacattgtgagcggtgacttctccctctcc 831
 QY 784 agcttactatggaagagacaggtgtgagctgacacttactatctgttgcctccaca 843
 Db 832 actcaccctcgagacacgataagattgtctgtctcttccatcactggtgtgacacccatg 891
 QY 844 ttaaacccgagtggtacagcttccagaaatagagagatgacggcagaaatgagaagt 902
 Db 892 ataaaccccttactctacagcttgagaaatagatgaaggtgacctgagaagact 950
 RESULT 14
 ABK16620
 ID ABK16620 standard; cDNA; 978 BP.
 AC
 XX ABK16620;
 DT 14-MAR-2002 (first entry)
 XX
 DE Human G-coupled receptor (GPCR) cDNA, seq ID No 29.
 XX
 KW Human; cytosolic; neuroprotective; immunosuppressant; mototropic;
 KW anti-inflammatory; anti-viral; gastrointestinal; cardiovascular;
 KW cerebroprotective; G-coupled receptor; cell proliferative disease;
 KW lymphoma; leukemia; breast cancer; cirrhosis; neurological disorder;
 KW stroke; Alzheimer's disease; multiple sclerosis; mental retardation;
 KW cardiovascular disease; atherosclerosis; angina pectoris; indigestion;
 KW congestive heart failure; gastrointestinal disorder; dysphagia; AIDS;
 KW gastritis; autoimmune disorder; inflammatory disorder; Crohn's disease;
 KW systemic lupus erythematosus; metabolic disorder; diabetes; obesity;
 KW viral infection; herpesvirus; parvovirus;
 KW acquired immune deficiency syndrome; ss.

OS Homo sapiens.
 XX
 PN WO200190359-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 22-MAY-2001; 2001WO-US16833.
 XX
 PR 22-MAY-2000; 2000US-206222P.
 PR 25-MAY-2000; 2000US-207446P.
 PR 02-JUN-2000; 2000US-208634P.
 PR 02-JUN-2000; 2000US-208661P.
 PR 07-JUN-2000; 2000US-209868P.
 XX
 PA (INCYTE) INCYTE GENOMICS INC.
 XX
 PI Patterson C, Tribouley CM, Yao MG, Griffin JA, Thornton M, Lu Y;
 PI Kallik DA, Gandhi AR, Au-Young J;
 XX
 DR WPI: 2002-106199/14.
 DR P-PSDB: AAU80498.
 XX
 PS Claim 5; Page 140; 148pp; English.
 XX
 CC The invention relates to a novel human G-coupled receptor (1). (I) and
 CC its corresponding polynucleotides are useful for diagnosing, treating or
 CC preventing cell proliferative diseases (e.g. lymphoma, leukemia, breast
 CC cancer or cirrhosis), neurological disorders (e.g. stroke, Alzheimer's
 CC disease, multiple sclerosis or mental retardation), cardiovascular
 CC diseases (e.g. atherosclerosis, angina pectoris or congestive heart
 CC failure), gastrointestinal disorders (e.g. dysphagia, indigestion or
 CC gastritis), autoimmune/inflammatory disorders (e.g. AIDS, Crohn's disease
 CC or systemic lupus erythematosus) or metabolic disorders (e.g. diabetes
 CC or obesity), or viral infections (e.g. infection by herpesvirus or
 CC parvovirus). ABK16615-ABK16637 represent novel human G-coupled
 CC receptor coding sequences of the invention.
 CC
 SQ Sequence 978 BP; 228 A; 258 C; 173 G; 319 T; 0 other;
 XX
 Query Match 27.7%; Score 255.8; DB 24; Length 978;
 Best Local Similarity 55.3%; Pred. No. 4,4e-69;
 Matches 497; Conservative 0; Mismatches 402; Indels 0; Gaps 0;
 QY 4 aatcaagctgtgaactagatcattatcttgggctccacaaagctgaacccag 63
 Db 52 aaccaaaccacatccgtaattcctccgggagtttccaagaagatgagatcaa 111
 QY 64 ggaattatctctcttcttctcattgtctatctgtgtcttcttcggacaatgctc 123
 Db 112 aacctctcttctgtcttcttctgtgtatgctcgtgacatgattgggaagcgctc 171
 QY 124 atcatcattgccaaatctatagcaacacctgtgacagcgcacatgatttctctctg 183
 Db 172 atcattgtgctgacgtctgatacgttccattcaccatgaccccatgattctctctgc 231
 QY 184 acactgctgtgtgacatctcgcacaaagacatcatccgaagatgtctgggccc 243
 Db 232 aatctactcttgctgatacttccatcttccaaactcagtcaccaaatgtctgtgaat 291
 QY 244 atgctacacacgaataatcattcattatgacagctgagctgacagctctctgttc 303
 Db 292 attcaaaccaagaatcaatccatctctatagagctgacacaaagatgatacttct 351
 QY 304 acatgtctctcggagctgagatggtctctccacacacatgagctatgacagctatg 363
 Db 352 atgtgttctgctgcatcacaattgctcttgaggacatgagctatgacacattgctg 411
 QY 364 gccattgttccctcttattacagatactatgataaacacacatggtgtgagctgtg 423

Dd		412	gcgactcggcaccccttcgaaattacaatctcatcgagcccagggttggcaatttgctc	471
Oy		424	ctcagacatgtaagcagtattgcagtcaccaatcttcgggtgacacagaactctaacatg	483
Dd		472	acgaatcatctaatggtctccctccagaaatattatgtgctctgacacacaccttcgtcat	531
Oy		484	aggttgacttcvtgtaggccaacacacatgacccaactctctcgtgtagatacccccatg	543
Dd		532	caatgctctctctttaaccacaaacactctccacaaactcttcctgtgacttvgccctc	591
Oy		544	ctgagcttgctcctlaagccctgtaagaatcaatgaggtgatggttatgtgtctgatat	603
Dd		552	ctcaaacgtgccttctcagataacttgtatcaagagcttggttgtttatgtgggttta	651
Oy		604	accctggccaatagggnacttatctctaactgcatctctcctaigtgtttatcatgttgt	663
Dd		652	tcaagtatacatctccctttaccctacgactcttcttcctatgctcgtgcatcataagct	711
Oy		664	attccgcgtatccgacagtagaaggccaaggagagcccttcctcaaatgcatcatc	723
Dd		712	gtcccgagagatcatcttcacacaaaggaaagtggaagcctctccacttggctctcac	771
Oy		724	ctcacagctgtagacccttactatctctcgttaactacacataatccgcctgtcttc	783
Dd		772	ctgacagctgtagtactcttctacgaaaccaatltgtagcgltgacttcttccctctcc	831
Oy		784	agctataattggaagagacaagctgtgtagctcactcatactctgtgactccacaa	843
Dd		832	actccacctgaggaactcgataagatttgctgtctccatcatcactgtggtgaaccccat	891
Oy		844	ttaaaccogatggtgtacagcttccagaataggagatgcaaggcagaataggaaagt	902
Dd		892	ataaacacctcatcatcagcttgtaggaataagatalgaagaagtlgccttgaanaagct	950
 RESULT_15				
AAAS1176	ID	AAA51176	standard; CDNA; 669 BP.	
XX	AC	AAAS1176;		
XX	DT	26-SEP-2000	(first entry)	
DE		Murine olfactory receptor ligand-binding region cdna 2.		
XX	NB	Degenerate primer; ligand-binding region; olfactory receptor;		
KW	NB	transmembrane domain; N-terminal; plasma membrane; translocation domain		
XX	XX	human rhodopsin receptor; odorant; toxicity; olfactory response; ss.		
OS		Mus musculus.		
XX	FH	Key	Location/Qualifiers	
FT	FT	CDS	1..669	
FT	FT	/product=	a Olfactory_receptor_ligand_binding_protein	
FT	FT	/partial		
XX	PN	WO200035274-A1.		
XX	PD	22-JUN-2000.		
XX	PF	17-DEC-1999;	99WO-US30221.	
XX	PR	17-DEC-1998;	98US-0112605.	
XX	PA	(UYJO) UNITV JOHNS HOPKINS SCHOOL MEDICINE.		
PI	XX	Reed RR, Krautwurst D, Yau KW,		
DR	XX	WPI: 2000-431471/37.		
DR	XX	P-PDB: AAY96663.		
XX				

PT primer pair for amplifying olfactory receptor nucleic acid, useful e.g.
PT for producing receptor libraries used in e.g. screening odorants for
PT toxicity

Claim 18; Page 30-31; 61pp; English.

XX
CC
CC
CC
CC
AAA51175-93 encode murine odorant/ligand binding regions of olfactory
receptors. They were amplified using degenerate primers shown in
AAA51163-64.

CC AA041053.04. CC
CC The primers can be used to generate a library of OR-LBR comprising the CC
CC transmembrane domains (TMD-III-VI, III-VI, III-VII or III-VII). Chimeric CC
CC nucleic acids encoding a 7-TMD protein containing a sequence encoding an CC
CC N-terminal plasma membrane translocation domain (especially the sequences CC
CC shown in AA966585), a first TMD and the OR-LBR. The translocation domain CC
CC was initially derived from the N-terminus of the human rhodopsin CC
CC receptor. LBR produced by amplification with the primers are useful for CC
CC generating new odorants, to screen for toxicity or therapeutic activity CC
CC in odorants, and altering an animal's olfactory response (claimed).

XX	Sequence
50	669 BP; 122 A; 202 C; 150 G; 195 T; 0 other;

Query Match 27.6%; Score 255; DB 21; Length 669;

Query Match:			
Best Local Similarity	61.5%	Pred. No. 6,3e-69;	
Matches 408; Conservative	0;	Mismatches 255;	Indels 0; Gaps 0.

187 ctggctgttgacatcatctgcacaacagcatcataccgaagtgtgggaccatg 246

Db 7 ctggccacatgacattgtgtgcaacctctgtgatccctaaggccctgatggccta 66

QY 247 ctaacatcagaataataccatttcatatgcagcgtgcattgccagctcttctgtcaca 300

Db 67 gtgtctgaagaacacccatctcttcaagatgcatggtcagctcttcttctg 120

307 **tggtcttgaggctgagatggttctctctcaccacatgacctatgtggcc** 360

Db 127 tgggtcctgtcttcggagctgtgtgtcctcaaggtcatgtgacctatgtgccc 180

QY 367 attgttccctcttattacagtactatatatgaaccacatatgtgtgtagcctgtc 420

Db 187 atctgtcttccctgcactacagctctagaatgagcccaagctctgtgtgggccctggcc 24

QY 427 agcatgtcatgtctattgcaagtcaccaattccctggtgcacacagctcttatcatgagg 48

Db 247 gtgggtgtatgtccatctgtgtctgtgaatgcatctgtgcacactggcctgatgacacgg 30

487 ttgacttctgtggccaacacattgaccaattcttctgtgagataccccattgctg 34

Db 307 ctgtcattctgtgcccccaagtcatcacccacttcttctgtgagatcccccacccccc 30

547 gcttgtcctgtagccctgtagaatcaatgtagtgatgtgtatgttgcgtgataacc

Db 367 ctgtcttcctgtagtcaccacatatcatatagcgttatgacacttgtggcagatgcctll 42

607 ctggcatagggacttattcttaacctgcattctcctatgggttcataccatcgcacac 88

Db 427 tatggtgcataacttgtgtctaacttgttatccatlggcctgcacatcggcagcgcc 48

667 ctccgtatccgcacagtagaaggcaagagggccttcctcaacacgctccaccacc / 2

Db 487 ctgcgcatacgcttctgtctgaggcagaagggcctllcclacccgllccalcccccccc

QY 727 acagtgtgacccttactattctcctgtaactacacccataaccgcccgcaccagc / 8

Db 547 atcgtgtctcagtgtaactatcatctgtgtctcgtgccaatgccaagaccgcgcccaagc

QY

787 tatacatltgaagacaaagtgtgtagcttcgcattccataccctcggcgaccccccaaca

Db 607 tacagccagaagaagcaagttaacctccgtcgtctgtaclcgalarccacgccccaaaccccg

QY	847	aac	849

Db 667 aac 669

Wed Jun 26 09:41:36 2002

us-09-975-308-8.rng

Page 15

Search completed: June 21, 2002, 12:29:40
Job time: 3865 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 10:16:45 ; Search time 49.43 Seconds
(without alignments)
4591.652 Million cell updates/sec

Title: US-09-975-308-8

Perfect score: 924

Sequence: 1 atgaatcacagcgttgaac.....ttgattcttgaacactact 924

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	199.8	21.6	900	US-09-085-371-5	Sequence 5, Appl1
2	189.6	20.5	966	US-08-748-506-7	Sequence 7, Appl1
3	184	19.9	966	US-08-748-506-8	Sequence 8, Appl1
4	182.8	19.8	966	US-08-748-506-5	Sequence 5, Appl1
5	182.2	19.7	1290	US-08-827-291A-1	Sequence 1, Appl1
6	177	19.2	966	US-08-748-506-6	Sequence 6, Appl1
7	174	18.8	1713	US-08-467-948A-1	Sequence 1, Appl1
8	174	18.8	1713	US-08-467-947A-1	Sequence 1, Appl1
9	149	16.1	984	US-08-748-506-9	Sequence 9, Appl1
10	94	10.2	1828	US-08-988-876-2	Sequence 2, Appl1
11	86.2	9.3	963	US-09-439-313-526	Sequence 526, App
12	83	9.0	1474	US-08-465-980-1	Sequence 1, Appl1
13	83	9.0	1474	US-09-053-303-1	Sequence 1, Appl1
14	83	9.0	1474	PCR-US95-07093-1	Sequence 1, Appl1
15	55	6.0	1320	US-08-599-252-84	Sequence 84, Appl1
16	55	6.0	1320	US-08-436-074-57	Sequence 57, Appl1
17	55	6.0	1320	PCR-US96-06352-84	Sequence 84, Appl1
18	55	6.0	1320	PCR-US96-06583-84	Sequence 84, Appl1
19	52	5.6	1506	US-07-937-608-13	Sequence 13, Appl1
20	52	5.6	1506	US-08-029-170-13	Sequence 13, Appl1
21	51.4	5.6	7218	US-08-232-463-14	Sequence 14, Appl1
22	46.2	5.0	1356	US-07-978-892A-4	Sequence 4, Appl1
23	46.2	5.0	1969	US-07-937-609-28	Sequence 28, Appl1
24	46.2	5.0	1969	US-08-029-170-28	Sequence 28, Appl1
25	45.6	4.9	1621	US-07-937-609-22	Sequence 22, Appl1
26	45.6	4.9	1621	US-08-029-170-22	Sequence 22, Appl1
27	45.6	4.9	1686	US-08-029-170-30	Sequence 30, Appl1

28	44.2	4.8	2243	1	US-07-937-609-15	Sequence 15, Appl1
29	44.2	4.8	2243	4	US-08-029-170-15	Sequence 15, Appl1
30	43	4.7	1440	1	US-07-978-892A-1	Sequence 1, Appl1
31	43	4.7	2015	1	US-07-937-609-25	Sequence 25, Appl1
32	43	4.7	2015	4	US-08-029-170-25	Sequence 25, Appl1
33	42.2	4.6	1209	5	PCR-US95-05616-1	Sequence 1, Appl1
34	41.6	4.5	936	2	US-08-288-663A-13	Sequence 13, Appl1
35	41.6	4.5	1194	2	US-08-288-663A-2	Sequence 2, Appl1
36	41.6	4.5	1228	2	US-08-288-663A-3	Sequence 3, Appl1
37	41.4	4.5	1260	1	US-07-866-979-3	Sequence 3, Appl1
38	41.4	4.5	1260	2	US-08-466-906B-3	Sequence 3, Appl1
39	41.4	4.5	1260	3	US-08-706-281A-3	Sequence 3, Appl1
40	41.4	4.5	1260	4	US-09-201-746-3	Sequence 3, Appl1
41	41.4	4.5	1260	4	US-09-097-231-3	Sequence 3, Appl1
42	41.4	4.5	1621	1	US-08-722-001-13	Sequence 13, Appl1
43	41.4	4.5	1776	1	US-08-722-001-29	Sequence 29, Appl1
44	41.4	4.5	2140	1	US-08-334-698-1	Sequence 1, Appl1
45	41.4	4.5	2140	1	US-08-228-932-1	Sequence 1, Appl1

ALIGNMENTS

```

RESULT 1
US-09-085-371-5
; Sequence 5, Application US/09085371
; Patent No. 6218358
; GENERAL INFORMATION:
; APPLICANT: Firestein, Stuart
; APPLICANT: Zhao, Haidong
; TITLE OF INVENTION: Functional Expression of, and Assay for, Functional Cellular
; TITLE OF INVENTION: Vivo
; FILE REFERENCE: P0151US2 / 09805059
; CURRENT APPLICATION NUMBER: US/09/085,371
; CURRENT FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: US 08/891,243
; PRIOR FILING DATE: 1997-07-10
; PRIOR APPLICATION NUMBER: US 60/045,961
; PRIOR FILING DATE: 1997-05-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Rat
US-09-085-371-5

Query Match          21.6%; Score 199.8; DB 4; Length 900;
Best local Similarity 52.9%; Pred. No. 3.2e-53;
Matches 460; Conservative 0; Mismatches 397; Indels 12; Gaps 1;

QY 16 gtaactgaattcattatcttggtgctcaccacaaagcctgaactccagggaattatcttc 75
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 28 gtgaggaatttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 87
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 76 ccttttctcatctgtcattctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 135
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 ttccttctctctgtgacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 147
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 136 aaaaatcatagcaacacactgcatagcaatgtatgtttctcttgacacatgtgtgt 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 148 attaggaacacacacacacacacacacacacacacacacacacacacacacacacac 207
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 196 gtggacatcatctgcaacacacacacacacacacacacacacacacacacacacacac 243
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 ctggagattgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 267
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 244 atgctaacttcagaaatcacatcttcatatgtacagtgatgtgtgtgtgtgtgtgtgtgt 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 aaggagaacacatgtacagtgatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 327
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 304 acatgtctctctgtgagctgagatgtgtctcttccaccacacatgtgcatatgtacgtatgt 363

```

Db	328	ctgggcgttggtgcacaagagtgctgcctctctgtcgtgtaatgagccatgacccgcatg	387
Oy	364	gcaattgtttccctctcttcatacaagtactatgatgaacacacatagtgtgtagcctg	423
Db	388	gcatctgcataccatccatccatccaccgcgcatgtgtaagtagccggatgtgtgcagaag	447
Oy	424	ctcagcatggtcatgctatgtcagtcagtcacacaaattccctgtgtgcacacagccttatcatg	483
Db	448	gcacgttgatccctggcgtgtagaggtttgtgatatccatggttaagtatttcctattctc	507
Oy	484	aggttgactctctctgtggccacaacacccatggacactctctctgtgtagaaccoccatg	543
Db	508	cgcctcgtcttaactgtggccccaacacatcaacacatttctcgtgatgtgtctccatgt	567
Oy	544	ctggctcttccctgttagccctgttaagaatacgaagtgtgtgtatgtgtgtat	603
Db	568	ctcaaccgtgtatgcactgacatgaatgtccacagcagaagcttcaagaacttgtccctggccatt	627
Oy	604	aacctggccataaggaggaacttatctctacacgtcactgctcattgtttacatgttgct	663
Db	628	tttatctcgtcgtggagccgctctcgttaactgtgggcattccataatgycatccacagtgct	687
Oy	664	attctccgtatccgcacagtagaagaagcagaagaagcctcttcaaatgtctatccat	723
Db	688	gtgatgcgcatccctccagctcgtcgtgcgcgcgcataaagccttttaacatctgtccctccac	747
Oy	724	ctcacagtggtgaccccttatactctctctgtaatctacacactataccgcgcctgttcc	783
Db	748	ctcactgtttgtagatcatcttctatgacgcagatattttacatactatggcagcctaagca	807
Oy	784	agctatacatcttgaagaagaagaagtgtagctgcactctatactctctgtgactccaca	843
Db	808	ctctcagcttttgacacacaagaagctgtgctctgtactctacgcgtgtcaattgttgcagtg	867
Oy	844	ttaaacccgatgtgtgtacagcttccagaa	872
Db	868	tccaatccatcatcatctactgtcttgcgaa	896

RESULT 12
 US-08-748-506-7
 Sequence 7, Application US/08748506
 Patent No. 6159707
 GENERAL INFORMATION:
 APPLICANT: Romnett et al.
 TITLE OF INVENTION: NOVEL SPERM RECEPTORS
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Leydix, Voigt & Mayer, Ltd.
 STREET: Two Prudential Plaza, Suite 4900
 CITY: Chicago
 STATE: IL
 COUNTRY: US
 ZIP: 60601-6780
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn release #1.0, Version #1.25
 CURRENT APPLICATION DATA: US/08/748,506
 APPLICATION NUMBER: US/08-NOV-1996
 FILING DATE: 08-NOV-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/033,751
 FILING DATE: 09-NOV-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 REFERENCE/DOCKET NUMBER: 74940
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-616-5600
 TELEFAX: 312-616-5700

```

: INFORMATION FOR SEQ ID NO: 7 :
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: type: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-748-506-7

```

Query Match	20.5%	Score 189.6	DB 3	Length 966
Best Local Similarity	51.8%	Pred. No. 5.3e-50		
Matches 429		Mismatches 399	Indels 0	Gaps 0

OY	71	tcctccctcttttctcatgtgcatctctgtgctttctctgcgaacagtgctacatca	130
Db	98	tgtttacaccttattcttctctatgtttcttattactaacaggaattactctatgacc	157
OY	131	ttgccaaatctctatagaacaccttgatacgcccatglatgtttctctctgcacatg	190
Db	158	tttgcatTTTGTACAGTCCATCTCTACACACCCCAAGTACTCTTCTTGCGCCACTGT	217
OY	191	ctgtgtggacatcatctgcacaacagcctctaccgaagatgctcggggaccatgtaa	250
Db	218	CTCTCTGTGAGATTGGCTATCTCTGTGTCATGTCATACCCAGAAAGTCTGACAGCCTTGTA	277
OY	251	catcgaataatcaccttcatcatgcagctgcacatgctccagcctctcttgltacatgt	310
Db	278	GTGAGGCCCGAAGGATCTCTTGGGAGGGTGTGTGCTACAGAGTCTCTTCTCATATTC	337
OY	311	ctctgagagctgagatggttctcttcacacacatgacctatgacgctatgtygcatt	370
Db	338	TTGTGATTAAGTATGAGTGTGCTATTTGGCAGCATGCGCTTTGACCGCTAATGTGGCTAT	397
OY	371	gtttccctcttattacagatctattatgaacacacatatgtgtgtgactgtctgaaga	430
Db	398	GTTCGCCATCTCCATATGCACACCCGAAAGATCTCGTGGGATATGTGCTATTGGCAATTG	457
OY	431	tgttcatgtctatctgcagtcacaaattctctggtgtgacacagctctctatcagaattga	490
Db	458	TCTCATAGGGTGATGGGATGATAGTAGTGTGCGACAGACCAATTTAAATTTCTCTGTGA	517
OY	491	ctttctgtggcccaaacacattgccaactctctctgtgagatacccccatltgtgct	550
Db	518	ACTTCTGTGTGACCCCTGTGATAGTACCACTTCTCTGTGACCTTCCACCTCTCCTGGCAC	577
OY	551	tgtccgttagccctgtlaagaatcctaagtagtattgtatgtgtctgataatcaacctg	610
Db	578	TTTCCCTGTGTATATATATCCCAAAATGAGGGTGCATCTTGTGGACAGATGCTTGCA	637
OY	611	ccatagggagacttattcttaccctgcacatcctatagttattatcatgtgtctatctcc	670
Db	638	TATTTAGTCCATTATTACTGATCATTTTCTCTATGTAGAAATTCCTCTCAAGTCTG	697
OY	671	gtatccgaagaagaaaggaagcaagagaagcctcttcaaatgtcatctcatatccaag	730
Db	698	TGATGCTTACCTGTAGGAGGGGCCCAATAAGCTCTCTCAACGTTCACTCACTCACTTGG	757
OY	731	tgtgtgaccttactattctctgtatactatacctatactccgcctctccagctata	790
Db	758	TATGTCACACCTCTTATGTGGCTCAACATCTGCCACACTATTGTGAGTCACAGTCTACCCACT	817
OY	791	catltgaagaagacaaggtgtgtagctgcacatcatatctctgtgactcccaatlaaac	850
Db	818	CACACAGAGAGTGACAAACTTTTGCCCTCTTCTATATCATCACTGACATTCACGCTGATTC	877
OY	851	cgatggtgtacagcttcacgaatcaggaagatgcaagtcagaagaaattaga	898
Db	878	CCATCATCTACAGCTTAAAGACACAGGAAGTAAGGGTGCACCTGAGAA	925

RESULT 3
US-08-748-506-8

Sequence 8, Application US/08748506
Patent No. 6159707

GENERAL INFORMATION:
APPLICANT: Ronnett et al.
TITLE OF INVENTION: NOVEL SPERM RECEPTORS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,506
FILING DATE: 08-NOV-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,751
FILING DATE: 09-NOV-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 74940
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5600
TELEFAX: 312-616-5700

INFORMATION FOR SEQ ID NO. 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-748-506-8

Query Match	Similarity	Score	ID	Length
Best Local Similarity	51.1%	Pred. No. 3.1e-48		
Matches	433	Conservative	0	Mismatches 415; Indels 0; Gaps 0
QY	71	tcctctctcttttttcacatgctacatctctgtgtgtcttcttcggaacaatgctacatca	130	
Db	98	TGTTACCCCTCATCTTCTCATGTCTTATTAGTATCACTAACGGAATAGCTCTCATAGGCC	157	
QY	131	ttgcgaaaatctatatagcgaacaccttgcatagcgccatgtatgtttctcttcgaacatg	190	
Db	158	TTTGCAATTTGTACCGAGTCCATCTTACACACCCCATGTACTCTTCTTGGGCCAATCT	217	
QY	191	ctgtgtgtgacatactcgcgaacaagaatcatcagaagatgtctggtggaacatgctaa	250	
Db	218	CTCTCTCTGGAGTTGGCTATACTTCTCTGTCTATACCAAGATGCTGGCAGATCTTGTA	277	
QY	251	cacgaagaataacacattcatatagctcagcgtgcgaatgccagctctcttctgttcaaatg	310	
Db	278	GTGAGGCCGAGAGATCTTTCAGGTGGGAGTGTGCCACAGATGTTTTCTTCATATCT	337	
QY	311	ctctcggagcgcagagatgcttctcttcacacacatggcctatgcagcgcatactggtgcat	370	
Db	338	TTGGATATACGTAGTGTCTGCTTATTTGGACAGCCATG6CCTTTAACCGCTATATGGCTAT	397	
QY	371	gttctcctcttcatacagtaactataltalagaacacacatatagtgtgtagcctgtctoa	430	
Db	398	GTTCGCCACTCCACATATGCAACCCCAATGAGTCGTGAGGTATGTGCCACTGGCAATG	457	
QY	431	tggtcatatgcatatgcagtaacaaatcttcctggtgtgacacagcctttatcagagattga	490	
Db	458	TTTCATGGGTGATGGGATGATAGTAGCTGTGGGACAGCAATTTATTTCTCTCTTA	517	

Oy	491	ctctctgtgaggccaaacacccattgacacactctctctgtgagatcaccccaattgctgctt	550
Db	518	ACTTCTGTGGACCCCTGTGTAGATAGAACCACTTCTTCTGTGATCTTCCACCTCTCTGGCAC	577
Oy	551	tgtccctgtagaccctgttaagaatcaatgaggtgtagtgtatgltgtcgtatataacccttg	610
Db	578	TTGGCTGTGGTATACATCCCAATTGTAGCGCTGCACATCTTGTGTGGTATGTCTCTTGCA	637
Oy	611	ccataggggaacttattcttaccctgcacactctcatggtttatcatgtgtcattctcc	670
Db	638	TATCTAGCCCTTTTTCGTGATCATTTAATCTTATGTGAGAAATCTCGTTGCAGGTGTGG	697
Oy	671	gtatccgcacagtagaagcgcaaggaagcgctctcaatgctcaatgccatccacg	730
Db	698	TGATGCCCTTACACGTGAGGGGGCCACAAGGCCCTTTCACTGTCTCTCCACCTACTTG	757
Oy	731	tgtgtgacccttaactatctctctgttaaltctaacctatataatccgacctgtctccagcata	780
Db	758	TAGTCACACCTTTTTATGTGGCTACAGGATCTGTACTCTTTTGGGCTCTAGCTTAGCCACT	817
Oy	791	catctgaaagagcaaggttgtagctgcaactctactctctgtgtgaccccaacttaac	850
Db	818	CACCAGGAATGGACAAACTCTTGGCCCTCTTTCACACGCAAGTATACATTCATGTGTGACC	877
Oy	851	cgatgtgtgtacagctctccagataatggagatgcatgcagcaagaatlagaaagtglttgcat	910
Db	878	CTATCATCTTATATGTTTAAAGAAACAAGGATGTCAAGCGAGCACTGAGAAGAATCTGGCCC	937
Oy	911	tcttgaaatg	918
Db	938	TGAATAAA	945

RESULT 4
 US-08-748-506-5
 Sequence 5, Application US/08748506
 Patent No. 6159707
 GENERAL INFORMATION:
 APPLICANT: Ronnett et al.
 TITLE OF INVENTION: NOVEL SPERM RECEPTORS
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Leydig, Volt & Mayer, Ltd.
 STREET: Two Prudential Plaza, Suite 4900
 CITY: Chicago
 STATE: IL
 COUNTRY: US
 ZIP: 60601-6780
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/748,506
 FILING DATE: 08-NOV-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/033,751
 FILING DATE: 09-NOV-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 REFERENCE/DOCKET NUMBER: 74940
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-616-5600
 TELEFAX: 312-616-5700
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 966 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)

US-08-748-506-5

Query Match	19.8%	Score 183.87	DB 3	Length 966
Similarity	51.2%	Pred. No. 7.4e-48		
Best Local				
Matches 424	Conservative	0	Mismatches 404	Indels 0
				Gaps 0

OY	71	tcttcctctttttctcaatgtctatctctgttgcttcttccttcggcaacatgcacatca	130
Db	98	TGTTCAACCTTCATCTCTTCATGTCTTATATACATAACAGAAATACTCTCATAGTCC	157
OY	131	ttggccaatactatagaacaaccttgcatcggcca tga tga ttttctctcttgacac tgg	190
Db	158	TTTGCAATTTGACAGTCCATCTCTACACCCCAAGTACTCTTCTTGGCCAACTGT	217
OY	191	ctgtgtgagacatcatctgcacaacaagcatcataccggaagatgtcggggaccatgctaa	250
D	218	CTCTCTCTGGAGATTTGGCTATCTCTGCTCTGTCTCATACCCAAAGATGTGCAGAGCCTTGTGA	277
OY	251	catcagaanaataccatttcatatcagtcgagctgctgtccagcctctctctgttcaatggt	310
Db	278	GTGAGGCCAGAGAGATCTCTGGGAGGAGTGTGCCACAGAGTTTTTTTTTCGGCATTTT	337
OY	311	ctctcgagacgtgagatggttctctcttccacaacatgacctatgacagctatggtccatt	370
Db	338	TTGGATATAACTGATGCTGCTCATTTGGAGCCATAGGCTTTTGACCGCTGATGGCCATAT	397
OY	371	gtttccctcttcaatacagtatctattatgaacacacatgagtgtgtagctctgacaga	430
Db	398	GCCTCCCATCTCCATATGCACACCCGAATGATCTCGNAGATATGTGCCCATTTTGGCAATTTG	457
OY	431	tgtctatgtctatctgcagtcacaaatllctctggtgcacaacagctctcatatgaagtgtga	490
Db	458	TTTTCATAGGGGAATGGGATGCGATAGTAATCTGTGGCAACCAATTTATTTTCTCCCTGA	517
OY	491	cttctctgtggccaacaacacatctgacccctctctctctgtgagataccccaattgctgctt	550
Db	518	ACTTCTGTGTGACCCCTGTGAATAATGACCACTCTCTGTGACCTTCCACCTCTCCTGGCAC	577
OY	551	tgtcctgtgagccctgtgaagaatcaatctgagtgatgtgtglatgtgtcatataaecttgg	610
Db	578	TTTGCTGTGTGAGATTAATCCCAAAAGAGGGCGCCATCTTGTGTGATGACATCTCTTGCA	637
OY	611	ccataggggaacttattcttaccctgcatcctctatagtttataatgttgcattctcc	670
Db	638	TATCTAGCCCAATTTTTCGTGATCATTTATTTATGTCAAAATTTCTCATTTCACTGCTTC	697
OY	671	gtatccgacaagatagaagaagaagagagccctctcaacatgctcatcatcatcaag	730
Db	698	TGATGCTTTCACCTGAGGGGCGCCATAAAGCTCTTTCCACCGTTTCGCTCACCTACTTG	757
OY	731	tgtgtgaccttactatcttctctgtatcatctaacctatactccgcctgtctccagctata	790
Db	758	TAGTCACACATTTTATGCTCAGACATGATTAACATTTTGGAGGCCAAGCTTTACGCCACT	817
OY	791	catttgaagagacaagtggtgtgctgcgaactatatactcttctgtgctcccaataaacc	850
Db	818	CACACAGGAATGTGCAAAATTTTGCCCTCTTCTACACAGTAACTGACATCAGCTGAACC	877
OY	851	cgatgtgtgtacagcttccaaatagagatgtcaggcagaagaaatlaaga	898
Db	878	CTATCATCTTATAGTTTAAAGAACAGAGAAATCAAGCAGACACTGAGAA	925

```

1 RESULT 5
2 US-08-827-291A-1
3 ; Sequence 1, Application US/08827291A
4 ; Patent No. 5874243
5 ;
6 ; GENERAL INFORMATION:
7 ;
8 ; APPLICANT: MacInn, Roberto
9 ;
10 ; APPLICANT: Sathe, Ganesh
11 ;
12 ; TITLE OF INVENTION: NOVEL OLRC15 RECEPTOR
13 ;
14 ; NUMBER OF SEQUENCES: 2

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY:
ZIP: 19406
COMMENCE DEADLINE FORM:

COMPUTER NAME: KINGS
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastISO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927.291A
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: King, William T
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: GP500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5015
TELEFAX: 610-270-5090

Query Match	19.7%	Score 182.2;	DB 2;	Length 1290;
Best Local Similarity	50.1%;	Pred. No. 1.3e-47;		
Matches 454;	Conservative	0;	Mismatches 453;	Indels 0;
				Gaps 0;

OY	3	gaatacaagcgtgtgaactgaagtcattgaattatctatgtcgggcccacacaaaagccgaactca	62
Db	307	GAATTGCGACTTTCACACTCCGACCTTCACTTCTCTGGGAATTTCAATACAGGCCACCA	366
OY	63	gggaattactctccctctttttctcattgtgtctactgttgcttctcggacaatgct	122
Db	367	CACCTTCTCTTCTTTTCTGGGCTCGGCACATTTTTCAGTGGCCTTATGGAAGACTGT	426
OY	123	catcatcatctggcaaaatctctagcaaacacctggaagccgatgatgtttctctct	182
Db	427	CATGGTTTCCCTACATCACTGGAACCCACAGCTCCACACCCCATGTACTCTCTCTAG	486
OY	183	gagcatgtgctgtgtgtgagacatcatctgacacaaagaacatcatocgaagatcttggagc	242
Db	487	CCAACTGTCCTCAATGAGACCTCATGCTCATGTGCACACCGTACCCAGATGGCCTTCAA	546
OY	243	catgtacaatcagaanaataaccattctatagaggggtctgatgtcccaagctctctgtt	302
Db	547	CTACCTGTCCTGGACACCAAGTCCATTTCTATGCTGGTGTGGCCACACAAATTTTCTCTTA	606
OY	303	caaatgtgtctctggagagctgagatgtgtctctcttcacacaatgagctatgacogctatgt	362
Db	607	TACATACCTGCTTGGCTCGAATGCTTTCTTTTGGCTTTATGTGCTTATGACGCTACAC	666
OY	363	ggaccattgtttccctctctcatltaagaagtaacataltatagaacacacatgtgtgtgaacct	422
Db	667	TGCCAATTTGCCACACCCCTTAAGATACCAACATCTCATGAGCCCTTAATTTGTGGACTTAT	726
OY	423	gtccagcatgtgtcatgtgctatgtgcagtcaccaactctcgggtgtgaacacagctcttcatc	482
Db	727	GACTGCTTTTCTTGATTCCTGTGGGCTCTACAGATGAATCAATTTATGCTGTGGCCACATT	786

TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1713 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: both
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 116..1003
: US-08-467-947A-1

Query Match 18.8%; Score 174; DB 3; Length 1713;
Best Local Similarity 51.0%; Pred. No. 5.9e-45;
Matches 411; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

QY 4 aatcaacagctgttaacagtgatctatattctgggcccacacaaagccctgaactcag 63
DB 137 AATCAGACAAATGGTCACAGAGTTCTCTACTGGGATTTCTCTGGCCCAAGATTCA 196
Y 64 ggaattactctctcttcttctcattgctatctgtgtgtcttctcggacaatgctc 123
DB 197 ATGCTCCTCTTGGGCTCTTCCCTGTTCTATGTTCTCACCCCTGCTGGGAATGGGAG 256
QY 124 atcatcatgcccacaaatctataagaacaccccttgatcccaatgatagtttcctctg 183
DB 257 ATCTTGGGGCTCATCTCACTGGACTCCAGACTCCACACCCCACTGCTTCTCTCA 316
QY 184 acactgtgtgtgtgtgagacatctgacacaaagacatcacacgaatgtctggagac 243
DB 317 CACCTGGCCGCTGTCACATGCTATGCTGCAACACAGATGCCCAATGCTGTGAC 376
QY 244 atgttaacatagaanaataacattcatatgcaagctgcatgctccactctctgttc 303
DB 377 CTCTGATCCAGCCCAAGCCATCTCTTGTGCTGATGACACTGACTTCTCTT 436
QY 304 acatggtctctggagctgagatggttctcttcaacacacaggtctatgacgcctatgt 363
DB 437 TTGAGTTTGGACATACATGCAATGCTCTGTTGCTGATGCTGATGCTGATGCTG 496
QY 364 gccattgttccctctctcatcaagtaactatattgaacacacatgtygtgtagcttg 423
DB 497 GCCATCTGCCACCTCTCTCCGATTTCTATCATCATGACCTGGAAAGTGTGATCATCT 556
QY 424 ctcaagcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 483
DB 557 GGCATCACTTCTGACATGATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 616
Y 484 aggttgaattctctgtggtgccaacacacattgaacacattctctctgtgagatacccccat 543
DB 617 AGACTGCCCCCTTTTGTGGGCTGCTGAAATCAACCACTTCTTGTGAATCTGCTGCT 676
QY 544 ctgagcttgcctgtgagccctgtaagaatcatgagtgatgagtgatgagtgatgagtgat 603
DB 677 CTGAGGCTGCGCTGTGCTGATACCTGCTCAACCAAGGTGCTGATCTTGAAGCTGCTG 736
QY 604 accctggccataaggagacttattcttaacctgcatctccatagtttattatcatgttgc 663
DB 737 TTTCATCTCTGCTGGGACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 796
QY 664 atttcgcatcgcgaacagtgtaaggcaaggcaaggccctcacaatgctcatatctat 723
DB 797 ATCTTGAGGATCCAGTCTGGGGAGGCGGAGAAAGGCTTCTCACTGCTCTCCAC 856
QY 724 ctcaacagtgatgaccccttactatctctcttaactcaacacacatctatcgccgtctgc 783
DB 857 CTGCGGAGTAGGAGACTTCTTGGGAGGCGGACCACTGCTGATGATGATGATGATGATG 916
QY 784 agctatcatattgaagagacaagt 809
DB 917 CGCATCTCTGAGGAGACAGCAAGAGT 942

RESULT 9
US-08-748-506-9
: Sequence 9, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnett et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/748,506
: FILING DATE: 08-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/033,751
: FILING DATE: 09-NOV-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION: 74940
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5600
: TELEFAX: 312-616-5700
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 984 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-748-506-9

Query Match 16.1%; Score 149; DB 3; Length 984;
Best Local Similarity 47.8%; Pred. No. 3.3e-37;
Matches 431; Conservative 0; Mismatches 470; Indels 0; Gaps 0;

QY 1 atgaatacagaagctgttgaactgagttcatattctcggccctacacaaagccctgaactc 60
DB 28 AGAATGGAGACTTGTGCTCTGAGTTCACTTGAAGGGTACCTGTGGCCGAGCACTG 87
QY 61 caggaaattactctctcttcttctcattgtctatctctgtgcttctctcggacaatg 120
DB 88 AAGATCTCTTCTTCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 147
QY 121 ctcatcatcattgccaanaatctatagaacaacattgcatagccatgatagttctctc 180
DB 148 CTCATATATTACCATCACTGCTGAGACCAAGACGACGACGACGACGACGACGACG 207
QY 181 ctgaacagctgtgtgtgagatctatctgacacaaagacatcatcggaagatgtcgg 240
DB 208 AGACACTTCTCTTGTGTGAGGTGTTTATTAATCACTGCTATCCCAAGCTCTCAAC 267
QY 241 accatgtcaacatcagaanaataacattcatatagaagctgcatgtlcccaagctctctg 300
DB 268 ATCATTTCTGTAGGAGCAAAAGATTCCTTGGGCTGCTGCTGCTGCTGCTGCTGCTG 327
QY 301 ttcaatggtctctgagagctgagatggttctctcaccacacagctcatagaccgtat 360
DB 328 TATCTTGTGCTGGGGCAACAGGTTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 387

```

Oy 361 gtyggcaattgttccctctctcattacagttactatgaacacacattgtgttagcc 422
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 CTGGCCATCTGCAGAACTCTACATTTATTCACACCATATGAGCCCAAGATGCTTCCT 447
Oy 421 ttgtcgaatgtaacagtcattgcagtcagtcaccaatccctgggtgacacagctcttacc 480
Db 448 CTCCTTACTGTCTGTTATTTTGGGGTTCCTCTCATGGCCAGTCAGTGTGATGCTT 507
Oy 481 atgaggttagctctgtgtggccaacacacattgacacactctctctgttgagataaccocca 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 TCCAAAGACATTTTACTGTGGCCAAACATTAATTCCTACATTTTCTGTGATTTGGACCA 567
Oy 541 ttgttgctctgttcctgttagccctgtgaataatcaatgagtgatgtgtatgtgtcgtgat 600
Db 568 CTGGCAAACTCTCTCTGTTCCAGAAACCAAGCTATTTAGATGCGTGTTTTACCTGCT 627
Oy 601 attacccttgacatagaggagacattatctcttacctgcactctccctatggtttatcatgtt 666
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 GTATTGTGCTTTTGGCTTCCTCTCTTAATACCATCTTTGGATACAGCAATAATAGTATGTC 687
Oy 661 gctattctcgatattccgacagctagaagagagagagagagagagagagagagagagagagag 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 ACCATAGGAACTCCCTTCACGCCAGGAGGAGAGAGAGCTTTTTCACCTGCTCTCT 747
Oy 721 catctcaagctgtgagacccttactatctctcgtlaactctaacactataccgcccgtct 784
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 748 CATCTCATTTGCTCTCTCTCTAATGATGAGCAGCTGTCATTTAATATACCTGAAGCCAAAG 807
Oy 761 tccagctatcatttgaagaagagaaagtgtagctgacactctactctactgtgtactccc 840
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 808 CAGAGAACAGAGTGTCAGCAACCAACAGAGAGGCTGCTCTTGTGACATGTGTTGACACCC 867
Oy 841 acattaaacccgatgtgtacagcttccagataagagagatgagagagagagagagagagagag 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 866 CTTCGAACTCTGTCATCTACACCTCTGGCAACAGCAGTCCACCAAGGCTCTCAGGAT 922
Oy 901 g 901
   |
Db 928 G 928

RESULT 10
US-08-988-876-2
; sequence 2, Application US/08988876
; Patent No. 6063596
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED
TITLE OF INVENTION: WITH IMMUNE RESPONSE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,876
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

```

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0441 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4106
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1828 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOTO1
CLONE: 364702
US-08-988-876-2

Query Match	10.2%	Score 94	DB 3	Length 1828
Best Local Similarity	48.2%	Pred. No. 9.6e-20		
Matches	356	Conservative	0	Mismatches 375; Indels 7; Gaps 3
Oy	3 gaatcagcgctgttaacacgaggttaattcattccctgggcccacacaaagcctgaactcoa	62		
Db	792 GAACTCAAGAGATGTTCTAATTTCTCTCTCCTCCTGAAGAGTCTCAGGGGATCCAGAACTGCA	851		
Oy	63 gggaattaccttcctctctttttctcattgctatcctttgtgctttcttcggcaacttgt	122		
Db	852 GCCAGTCCTTGCGGGGCTGTCTCTGCATAGTGCCTGGTGACCGGTCTGGGGAACCTGCT	911		
Oy	123 catcatcattgcacaaaattatagaacacctgacgaagccgatgtatgtttctctct	182		
Db	912 CATTCAATCCGGGCATCAGCCCTAGCTCCACCTCCACCTCCACACCCCATGTACTCTTCTCTC	971		
Oy	183 gaacacggcctgctgtgtagacatactgcacaaagaacatcataccgaagatgctgggac	242		
Db	972 CAACCTGTCCTTGCTGCTACATCGGTTTCACCTCCOACACAGGTCGCCCAAGATGATTTGGA	1031		
Oy	243 catgttaacatcagaaaaataccaatttaatatgaagctgtgattccagctcttcgtt	302		
Db	1032 CATCCAGTCTCACAAGCAGAGATCATCTTCCTATGACGGCTGCTGACTGATGATGTCCTCTT	1091		
Oy	303 caaatgtctctcggaactgagatg----gttccttcacacaacatgacatgaacgct	358		
Db	1092 TGCCATTTTGGAGAGCATGGAAGAGACATGCTCCTGAGTGTGATGGCCATGACTGGT	1151		
Oy	359 atgtggccatttgttccctctcattacagtaactataaigaaccacatatgtgttag	418		
Db	1152 TTGTAGCCATCTGTCACCCCGGTATATCATCACCATCATGAACCCGTTTCTGTGCTT	1211		
Oy	419 ccttgctcaagcattgtaatgtgctatttgagtcac--aatccgggtgcacagacct	476		
Db	1212 TCTAGTTTGTGTCCTTTTCTTTTCTCAGTCCTTTAGACTCCCAAGCTGCAACTTGAT	1271		
Oy	477 tataatgaggttgcttctcgtctgtagccctgtgaagaatcaatgaggtgagtgtgttc	536		
Db	1272 TGCCCTTACAAAGTCCCTGCTTCAAGAGATGTGGAATTCCTAATTTCTCTGTGACCCCTC	1331		
Oy	537 cccattgtgctctgtctcgtctgtagccctgtgaagaatcaatgaggtgagtgtgttc	596		
Db	1332 TCAACTCTCCCATCTGTGATGTGAGACACTTCACATTTAAACAATTAACATGATTTTCC	1391		
Oy	597 tgataataccctgccaataagggaacttatacttaaccgcatctcctatggtttatcat	656		
Db	1392 TGCTGCCATATTTGGTTTCTTTCTCCATCTCAGGGACCTTTTCTCTTACTGTAATAAATCT	1451		
Oy	657 tgttgctattccgtctatccgacagtagaagccaagaggaagccttctcaaatgctc	716		
Db	1452 TTCTCTCATCTCTGAGGGTTTCTATCATCAGTGTGGGAATATAA--CCTTCTCACCTGTGG	1510		
Oy	717 atctcatctcaagctggt 734			


```

OY      212 caaaagacatcataccgaagaigtggyggacatatgtcaactaactogaanaatacccttcat 271
           |||
Db       491 CCACATCCACAATCAGCTTAAGATTCCTTGCCCCCTTTGTGTTGAATTCCCAGAAGATTTAGCA 550
OY      272 atgaagcgtgcattgcocagctcttcgtttcacatgbtctctvgagagttagatggltc 331
           |||
Db       551 TTGAGGCCTTCTTACCAGAAATGTTCTTAAATCAAAGGCCOTCCACCACTAATAATCCACCA 610
OY      332 tcttacaccacatgatgccttatgatcgcatgttgccaatttgttcccccttcaaatgta 391
           |||
Db       611 TCGCGTGCGGCATAGCGCTTTGACCGTTATGTGGCCAATCGCACCCCACTGCGCCATGCTG 670
OY      392 ctatatgaaccaaccatatbtgtgtgtagcctgtcasgaatgtlcaatgtgtatltgacgtca 451
           |||
Db       671 CAGTGCTCAACAATACACTAGTAACAGCCACAGATTNGCATTCGTGGCTGTGGTCCGCGGATGCC 730
OY      452 ccaatlctctgtgtgacacacagcctcttatcatatgaagtgtgacctctgtgyggccaaca 511
           |||
Db       731 TCTTTTTTMOCCACTCCTCTCTGTAATGAAGGCGCTGCTTTGCACTCACTCAAGTTGCC 790
OY      512 ttgaccactctctctgtgaaatagcccocatitgtgtgcttgcttgctcgtiaaggaa 571
           |||
Db       791 TCTCGCACTCCTATTGTGTGTCACCAGAGANGTAATGAAGTTGGCCTAATGCAAGACTTTGC 850
OY      572 tcaatgagtgatbgtygtalatgttgcgtatattaaccccttgccalagygagacttattctta 631
           |||
Db       851 CCATATGTGTAATATAGTCCTTACTGCGCAATTCGTGTGTCATFGGCGCGAGCAATGAATGTCA 910
OY      632 ccgtacatcctaatagtttataatattgtgtctattcttcgtatccgaagagtagaaggca 691
           |||
Db       911 TCTCTTTCTCTATTCTTGTGTAATAACGAAGGTTGTGCAACTGCTTCCAAAGTAAGAAC 970
OY      692 agagaagagccttcaaatgatgcatactatcatatcaacagtgtygacctttaattctc 751
           |||
Db       971 GGGCCAAAGCCTTTGGAAACCTGTGTGTCAACAAATGTGTGGTAATCGGCTCTATTATGTC 1030
OY      752 ctgtaatctaacct 766
           |||
Db       1031 CACTATTGGCCTCT 1045

RESULT 13
US-09-053-303-1
; Sequence 1, Application US/09053303
; Patent No. 5948890
GENERAL INFORMATION:
APPLICANT: Soppet, Daniel R.
APPLICANT: Li, Yi
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAU70
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
ADDRESSEE: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/053,303
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/465,980
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
```

```

NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
US-09-053-303-1

```

[illegible]

RESULT 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 09:13:22 ; Search time 1694.03 Seconds
(without alignments)
7361.845 Million cell updates/sec

Title: US-09-975-308-8

Perfect score: 924
Sequence: 1 atgatacagcgctgttaac.....ttgcatttcgaacactag 924

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gp_estl:*
10: gp_est2:*
11: gp_hlc:*
12: gp_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323.2	35.0	470	12	AQ428256 CITR1-E1-
2	200.6	21.7	2021	11	BC016940 Homo sapi
3	188.6	20.4	580	12	AZ593814 IM0405K03
4	175.4	19.0	3063	11	AK016560 Mus muscu
5	173.8	18.8	1394	11	AK017036 Mus muscu
6	173.2	18.7	726	12	AZ255734 RPCI-23-1
7	172	18.6	616	12	BH342053 CH230-51L
8	171	18.5	762	10	BC193339 RST12467
9	170	18.4	698	12	BH069789 RPCI-24-3
10	164.6	17.8	611	12	AZ103967 RPCI-23-3
11	159.2	17.2	853	12	BH331857 CH230-46N
12	159	17.2	679	12	AG167722 Pan trogl
13	157.4	17.0	642	12	AZ696227 2M0241J24
14	156.6	16.9	600	12	AQ976178 RST17016
15	155.6	16.8	796	10	BC197640 RST17016
16	154.4	16.7	842	12	AZ725747 RPCI-24-9
17	153.4	16.6	699	12	AZ555190 RPCI-23-1

18	153.2	16.6	1501	11	AK016338 Mus muscu
19	152.6	16.5	612	12	AZ396764 IM0161M22
20	152.4	16.5	805	12	BH272774 CH230-56L
21	150.2	16.3	629	12	AZ019257 RPCI-23-2
22	150	16.2	630	12	AQ503914 RPCI-11-2
23	149	16.1	605	12	AZ642411 IM0505L22
24	148.8	16.1	794	12	AZ913406 RPCI-24-1
25	147.8	16.0	627	12	AQ240757 CIT-HSP-2
26	147.8	16.0	680	12	AZ709687 RPCI-24-8
27	147.6	15.9	814	12	AZ699230 RPCI-23-2
28	147.2	15.9	686	12	AZ086625 RPCI-23-2
29	146.8	15.9	634	9	BB635510 BB635510
30	146.4	15.8	542	12	AZ252110 RPCI-23-4
31	146	15.8	743	12	BH322680 CH230-197
32	145.6	15.8	752	12	BH065531 RPCI-24-2
33	144	15.6	632	12	AZ765752 IM0562E19
34	143.8	15.6	542	12	AQ632709 RPCI-11-4
35	143.8	15.6	634	12	AZ507680 IM0349L20
36	143.2	15.5	635	12	AQ527378 RPCI-11-3
37	142.8	15.5	667	12	BH293359 CH230-92C
38	141	15.3	752	12	BH362510 CH230-127
39	140.2	15.2	545	12	AQ389818 RPCI11-14
40	140.2	15.2	632	12	AZ648626 IM0517B18
41	140.2	15.2	660	12	AZ380178 IM0136A04
42	140.2	15.2	732	10	BC6068751 H3068H10-
43	139.8	15.1	728	12	BH279965 CH230-38K
44	139.2	15.1	521	12	AZ407823 IM0178P09
45	138.6	15.0	594	12	AZ599283 IM0414H04

ALIGNMENTS

RESULT 1
LOCUS AQ428256 470 bp DNA linear GSS 24-MAR-1999
DEFINITION CITR1-E1-2578F11.TF CITR1-E1 Homo sapiens genomic clone 2578F11,
DNA sequence.
ACCESSION AQ428256
VERSION AQ428256.1 GI:4496022
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 470)
Zhuo,S., Adams,M.D., Niernan,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
AUTHORS Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CITR1-E1-2578F11.TF
Contact: Shaying, Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbest@igf.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tlgr.org/cdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES

source
1..470
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2578F11"
/clone_lib="CITR1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;

BASE COUNT 119 a 122 c 80 g 147 t 2 others
ORIGIN

Query Match 35.0%; Score 323.2; DB 12; Length 470;
Best Local Similarity 98.8%; Pred. No. 3.2e-72;
Matches 325; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgaacacagcgttgtaactagttacattatcttggcctcaacaaagcctgaactc 60
|||||
Db 142 ATGAATACAGAGCGTTGACTGATTCATATCTGGCGCTCAGCAAAAAGCTGAATC 201
QY 61 caagggaattact 120
|||||
Db 202 CAGGAAATATATCT 261
QY 121 ctatcatcatctgcaaaatctatagcaaaccttgcatacgccatgtatgtttctt 180
|||||
Db 262 CTGATCATCTGTCATATCTATACACACCTTGATGAGCCCAATGATGTTTCTT 321
QY 181 ctgacactggcgtgtgtgacatcatctgcacaaacagcttatccgaagatgctggg 240
|||||
Db 322 CTGACACTGGCGTGTGGACATCATCTGCAACAAGCATACCGAAGATGCTGGGG 381
QY 241 accatgctaatcagaagaataacattcatalgcaggtctgcattgtccagctctctg 300
|||||
Db 382 ACCATGCTACATCAGAAAATACCATTTCAATGACAGCTGCAATGCCAGCTCTCTTG 441
QY 301 ttccatgtctctcgggagctgagatgt 329
|||||
Db 442 TTCACATGTCCTGGGAGCTGAGATGCT 470

RESULT 2
BC016940 2021 bp mRNA linear HTC 09-NOV-2001
LOCUS Homo sapiens, similar to olfactory receptor, family 2, subfamily A,
DEFINITION member 4, clone IMAGE:4424116, mRNA.
ACCESSION BC016940
VERSION BC016940.1 GI:16877381
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2021)
Strausberg, R.
Direct Submission
Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mdedpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

REMARK
COMMENT
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 27 Row: 1 Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, Similarity but not identity
to protein!

This clone has the following problem: frame shifted.

FEATURES
Location/Qualifiers
source
1..2021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4424116"
/issue_type="Kidney, hypernephroma"
/clone_lib="NIH_MGC_89"
/lab_host="DH10B"
/note="Vector: PCMV-SPORE6"

BASE COUNT 470 a 546 c 427 g 578 t
ORIGIN

Query Match 21.7%; Score 200.6; DB 11; Length 2021;
Best Local Similarity 51.5%; Pred. No. 1.2e-40;
Matches 461; Conservative 0; Mismatches 434; Indels 0; Gaps 0;

QY 4 aatcacagcgttgtaactagttacattatcttggcctcaacaaagcctgaactcag 63
|||||
Db 438 AATCAGACAAATGGTCAAGAGATTCCTCTACCTGGGATTCCTGGGCCAAGATTCAG 497
QY 64 ggaattatc 123
|||||
Db 498 ATGCTCA 557
QY 124 atcatcatctgcaaaatctatagcaaaccttgcatacgccatgtatgtttctctg 183
|||||
Db 558 ATCTGGGGCTCATCTCACTGAGCTCCAGACCTCCACCCCATGATTTCTCTCTCA 617
QY 184 acactggcgtgtgtgacatcatctgcacaaacagcatcatcacgaagatgctgggagc 243
|||||
Db 618 CACCTGGCGGTGTCACATGCGCTTATGCTTCGACACAGTCCCGAGATGCTGTGAAAC 677
QY 244 atgctaacatcagaagaataacattcatalgcaggtctgcattgtccagctctctgtc 303
|||||
Db 678 CTCTCTGATTCACACCAAGCCCATCTCTCTGCTGCTGATCATAATGACCTTCTCTT 737
QY 304 acatgtctctggagctgagatgtctctctcaacacatgacctatgacctatgtg 363
|||||
Db 738 TTGAGTTTTCACATCTGATGATGCTCTGTTGTTGCTGATGCTCAGATGGATGAGTG 797
QY 364 gccattgttccctg 423
|||||
Db 798 GCCATCTGCCACCTCTCCGATATTCATCATCATGACCTGGAAATCTGCATCACTCTG 857
QY 424 ctgacatgtctatgtgctatgtcagtcacaaatctctgggtgagacagctctatcatg 483
|||||
Db 858 GCCATCACTTCTGAGATGTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 917
QY 484 aggttgacttctgtgggccaacaaacattgacacattctctctgtgagatacccatg 543
|||||
Db 918 AGACTGCCCTTTTGTGGGCTCTGGAATACACACACTTCTTGTGAATCTGCTGTGTC 977
QY 544 ctgacttctctctgtagccctgtaagaatcaagatgagtgatgtgtgtgtatatt 603
|||||
Db 978 CTCAGGCTGGCGCTGTGCTGATACCTGTGCTCAACAGAGTGATCTTTTGCAGGCTGATG 1037
QY 604 accctggcacaaggagacttattcttcaactgcatactcctcatggtttatcatgtgtct 663
|||||
Db 1038 TTCACTCTGTTGGGACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1097
QY 664 attctcgtatccgacagtagtaaggaaggaaggaaggaaggaaggaaggaaggaag 723
|||||
Db 1098 ATCTGAGGATCCAGTCTGAGGAGGCGCGAGAAAGGCTTCTCAGCTGCTCTCCAC 1157
QY 724 ctcaacagtgtagacccttatctctctgtaactacacattacacattacacacattcc 783
|||||
Db 1158 CTCTGCTGATGAGGACTCTCTTTTGGAGCGCATGCTGATGAGGCGCTGATGCTC 1217
QY 784 agctatcatattgaaagaaagaaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843
|||||
Db 1218 CGCATTCCTGAGGAGCAG 1277

Query Match	20.48;	Score 188.6;	DB 12;	Length 580;
Best Local Similarity	59.48;	Pred. No. 1e-37;		

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwada, K., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Komuro, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,

QY	843	atlaaccgcgaatggctcagcttcacagataggagaggttcgcagcgaattaggaagt	902
Db	1288	TGTGAACCCCTCATTTACAGCCTGAGAAATAGAGAGGATTGAGCTGTTAGAAGACT	1348
OY	903	gtttg 907	
Db	1349	ACTGG 1353	
RESULT	5		
LOCUS	AK017036	1394 bp	Linear
DEFINITION	AK017036	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933433E02; similar to T1 OLFACTORY RECEPTOR, full insert sequence.	HTC 19-JAN-2002
ACCESSION	AK017036		
VERSION	AK017036.1	GI:12856091	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Enayayota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3 (sites)		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4 (sites)		
AUTHORS	FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	5 (bases 1 to 1394)		
PUBMED	11076861		
REFERENCE	5 (bases 1 to 1394)		
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Ouehnbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, U., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUL-2000)		

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAGAAGAGATCCAGACTCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15'

GAGAGAGAGATTCTCGATTAAATTAAATTCACCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source

location/qualifiers

1..1394

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="MGD:MGI:1894709"

/db_xref="taxon:10090"

/clone="493343EB02"

/sex="male"

/tissue_type="testis"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

332..1282

/note="data source: SPTK, source key: 09EPF8, evidence: ISS putative similar to T1 OLFACTORY RECEPTOR"

/codon_start=1

/protein_id="BAB30564.1"

/db_xref="GI:12856092"

/translation="MEPWNSTLESGFLIVGLIDSSGPPELLCATVTLTYMLAISNGLLLLVITVDARLHPMYILLRQLSLIDLIFSVTPNTVPVLRLNDRTISEGALOLFSAMLTGGADELLAFMAYDRVAICHLPMIMSPACRLMAVISLISLALGHTVYTHMPFCMSOEIRHLCVEPLRKACADTSOELAMNVTVGVILFLPSLAITSYSLLFIETVLMHSPNEGRKKALVTGSLLTVGMFGAATFMVYLPSSPHSKQDNIIISVFYTIVPALNLAIYSLSLNKEVGAVRYVLRHLIPAHATV"

BASE COUNT 355 a 337 c 304 g 398 t

ORIGIN

Query Match 18.8% Score 173.8; DB 11; Length 1394;
Best Local Similarity 49.5%; Pred. No. 8.1e-34;
Matches 448; Conservative 0; Mismatches 457; Indels 0; Gaps 0;

Dy 3 gaatcacagcgttgtaactgaattcatattctgggacctccacaaagaacctgaattcca 62
|||||
Db 343 GAATTCATCACTGGAAATGGAATTCATCTGTGTGGGATTCGTGATGGCATACTGCCTGCC 402
|||||

Oy 63 gggaattacttcctccttttttctcaatgtctatctcttggtgcctttctcggaacaagtct 122
|||||

Db 403 TGAAGTCTCTGTGTGGCACAGTTTAACAACCTGTATCATGTGTGGCATTGATCAACATGGAGCT 462
|||||

Oy 123 catcatcatctgcacaaaatactatagacaacacctgicacagaccatcgtatgtttcctctc 182
|||||

Db 463 TCCTACTCCTGGTCAATCAACAGTGAATGCCGGCTTCACGTACCATTGAACCTCCATCTAG 522
|||||

Oy 183 gaacctggcctgttgtagacatcatctgcacaaagaacatcataccgaagatgctgggggac 242
|||||

Db 523 GCAGCTGCTCTTCATGACCTCCTCTTCACATCAAGTTGTAACTCCCAACACTGTTTGGGA 582
|||||

Oy 243 catgtctaaccatcagaanaatatcatattcatatgatcaggagctgcatagtgccagctctcttgt 302
|||||

[illegible]

FEATURES									
SOURCE									
Location/Qualifiers									
1..726									
/organism="Mus musculus"									
/strain="C57BL/6J"									
/db_xref="taxon:10090"									
/clone="RPCI-23-16619"									
/clone_id="RPCI-23"									
/sex="Female"									
/lab_host="DH10B"									
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1: ECORI; Site:2: EcorRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."									
BASE COUNT 167 a 171 c 118 g 270 t									
ORIGIN									
Query Match 18.7%; Score 173.2; DB 12; Length 726;									
Best Local Similarity 52.7%; Pred. No. 9,6e-34;									
Matches 376; Conservative 0; Mismatches 338; Indels 0; Gaps 0.									
63	tatctccctttttctcatatgctatcttcttgcttcccgcaaatgctcat	128							
Db	7 TCGTTCCTCTTGCTCTCTGTATATATATAGTACTAGACCGGAATTTGGT	66							
Qy	129 catgccaaatcatalagacaacacttgacatacgccatgtagtlttccctg	188							
Db	67 AATTCTAATTGGACGTGAATTCACCTTCACACCCGAGTACTTTTCTTTAACT	126							
Qy	189 gactctgttggacaatcatctgcgcaacaagcatcatcagcaagatgctgg	248							
Db	127 GCTTTGATAGACCTCTGTACTCTTCAAGTGTATACCCAAATCGTGAAC	186							
Qy	249 aacatcagaanaacacatctcatatgacagctgcatgctccatcttctg	308							
Db	187 ATTAATAAAGAAATATATATCTTATAGGGGGTATGACCCAACTATTTTAT	246							
Qy	309 gtctctggagactgagatggtctcttccaccacatggccttagacagc	368							
Db	247 CTTTGTTATTTTCAGAGTATGTATGTAAAGTAAAGGCTTAATGATAG	306							
Qy	369 ttgttccctctcatctacagtaactatagaaaccaacatagtgtagc	428							
Db	307 CTGTAATCCACTCTTATATATATTTGCTATGACACCTAAATATGTTCT	366							
Qy	429 catgtcatgagctatctgacgtacacaaattctctgggtggacaagc	488							
Db	367 TGGTTCATTAATGATGGCAATTTCTGGTGCATGCTACACAGATATG	426							
Qy	489 gacttctgtggcgcaacaacacatctgcacctctctgtgagatgaccc	548							
Db	427 GACCTTCTGTGATGCAACACACATCACTACTTCTGTGACATCTCCTCG	486							
Qy	549 ttgtccctgtgacccctgtlaagaatcaatggagtgtagtgtagatata	608							
Db	487 GCTTCCTCGACCGACGACCTATGTCAATGAACTGGAGGTTTCTTCG	546							
Qy	609 ggcataaggagacttatcttaccgtatcctcatggttttcaatgtgt	668							
Db	547 TATCCTGTGTGCCACGACATCAACCAATTTTATCTTTATGAGGTTCA	606							

Qy	669	ccgatccgcacaaagtcgaagcgaagcgaagcgccttcacacatgcataccatccac	728
Db	607	CCACATCAACCTCCCAATGAGGCGCAGGCCCTTCACACTGACACTGACAGTATCCCACTAAT	666
Qy	729	agtgatgaccccttactatctctcctgtaactacacataccatccgcgcctgttc	782
Db	667	TGCTGGTCTCTCTCTTCTTGCAATGAGGTGCATTTATGATCTTAAACCTCCTC	720
RESULT	7		
LOCUS	BH342053	616 bp	DNA linear GSS 03-DEC-2001
DEFINITION	CH230-51L13.1.vv CHORI-230 Segment 1 Rattus norvegicus genomic clone		
ACCESSION	BH342053		
VERSION	BH342053.1	GI:17272787	
KEYWORDS	GSS.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 616) Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.		
TITLE	Rat BAC End Sequences from Library CHORI-230 EcoRI segment		
JOURNAL	Unpublished (1999)		
COMMENT	Other_GSSs: CH230-51L13.1.vv Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/BACPac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@small.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/BACPac/or erling_Information.htm). BAC end pages: http://www.tigr.org/tdb/Bac_ends/rat/Bac_end_intro.html Plate: 51 row: L column: 13 Seq primer: T7 Class: BAC ends.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..616		
	/organism="Rattus norvegicus"		
	/strain="BN/SSNhsd/MCW"		
	/db_xref="taxon:10116"		
	/clone="CH230-51L13"		
	/clone_lib="CHORI-230 Segment 1"		
	/sex="Female"		
	/cell_type="Brain"		
	/note="Vector: pPARBAC2.1; Site_1: EcoRI; site_2: EcoRI; CHORI-230 Rat (BN/SSNhsd/MCW) BAC library produced by Pieter de Jong"		
BASE COUNT	142 a 152 c 108 g 214 t		
ORIGIN			
Query Match	18.6%; Score 172; DB 12; Length 616;		
Best Local Similarity	57.0%; Pred. No. 1.8e-33;		
Matches	333; Conservative 0; Mismatches 250; Indels 1; Gaps 1;		
Qy	164	ccatgatgtttctctcttcgtacacgcgcgtgtgtgagacatcatctgcacaaagcatca	223
Db	1	CAATGACTCTTTTCTCTTTTAACTTGTCATATTGTTGACCTCGGTGATCTTCAGTGTCA	60
Qy	224	taccgaagatcgttgaggaccatgcacacatcagaacatcacatctcatatgcaggctga	283
Db	61	CACCCAAATAGTGATGA-GTTTAAATATGAGAGAGACATTTATACCTACCAAGAGATGA	119

Oy 284 tttcccaactctcttgatccacaaatgctccggagagctgaagtggctctctccacaca 343
 Db 120 tgcacacagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 179
 Oy 344 tggcctatgacccgctatgcttgccatttgcttccctcttcatacagtaactatgaacc 403
 Db 180 tggcctatgacccgctatgcttgccatttgcttccctcttcatacagtaactatgaacc 239
 Oy 404 accataatgctgtagccttgctccagcaatgctatgctcatttgcaatccaaattccctgg 463
 Db 240 cgaatttattgtttgaacctgcgtcttgcctctacacataatggcattttctggctgcattg 299
 Oy 464 tgcacacagctcttcatacgaagtctgactctctctgggccaacacacattgacactctc 523
 Db 300 ctacacacagggcgatgcgtgagacatgactttctgtgatgcacaaacatctacactact 359
 Oy 524 tctgtgagatacccccaacttgcttgctcttgctctgtagccctgtaagaatcaatgaagtga 583
 Db 360 tctgtgagatacccccccttggcttcttcaactcttctgcacacgacctatgtgcataatgacgtg 419
 Oy 584 tgggtatgctgtagctatattaccctggcgcaatgaggaacttattcttaccgtcacctctc 643
 Db 420 aggtcttctgtgtttataggcattcaacatcttctgtgtcccgacattacacctcttcatctcct 479
 Oy 644 atggttctacatgcttgctcattctccglatccgcacagtagaaggaagaaggaagcct 703
 Db 480 atggttctacatcctccaagatatttttttcattatcagctcccaaggaaagcaggtccaaagcct 539
 Oy 704 tctcaacatgctcattcatctatctccacagatgggtgacacttaccat 747
 Db 540 ttcagcactgtccagttccacataatccgacagtttctctgtcttctt 583

	RESULT	8
BGI933339	LOCUS	762 bp mRNA linear EST 21-APR-2001
R5112467	DEFINITION	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
BGI93339	ACCESSION	BGI93339
BGI93339..1	VERSION	GI:13715026
EST.	KEYWORDS	
SOURCE		human.
ORGANISM		Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
1 (bases 1 to 762)		
Harrington,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,		
Cain,S., Leventhal,C., Thornton,C., Ramachandran,R., Whittington J.		
, Lerner,L., Costanzo,D., McElligott,K., Boozar,S., Mays,R., Smith		
,E., Veloso,N., Kilka,A., Hess,U., Colchen,K., Lo,K., Offenbacher		
'J., Danzig,J. and Ducar,M.		
Creation of genome-wide protein expression libraries using random	TITLE	
activation of gene expression	JOURNAL	
Nat. Biotechnol. 19 (5), 440-445 (2001)	MEDLINE	
21227151	COMMENT	
Contact: Scott J. Cain		
Athersys, Inc.		
3201 Carnegie Ave,		
Cleveland, OH 44115, USA		
Tel.: 216 431 9900		
Fax: 216 361 9596		
Email: scaine@atersys.com		
High quality sequence stop: 551.	FEATURES	
Location/Qualifiers	source	
1..762		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_lib="Athersys RAGE library"		
/cell_line="HT1080"		
/note="See 'Creation of Genome-wide Protein Expression		
Libraries using Random Activation of Gene Expression'		
Nature Biotechnology, In press. Note that even though the		
cult type indicated its HT1080, since a random activation		
method was used, these sequence tags are not necessarily		

BASE COUNT 152 a 253 c 156 g 201 t
 ORIGIN expressed in HT1080 under normal circumstances."

Query Match 18.5%; Score 171; DB 10; Length 762;
 Best Local Similarity 54.0%; Pred. No. 3.5e-33;
 Matches 394; Conservative 0; Mismatches 330; Indels 5; Gaps 2;

15 tgaatgagttatattatcttgagcctcaccacaaagcctgactgacaggaattatctt 74
 |||||
 39 TGCTCCGAGTTCTCTCTCTGAGATCTCCAGCGCCCGACAGACATCTCTCTT 98
 |||||
 75 cctcttttctcatctgctctatctgttggtcttctcggcaacagctcatctatcttc 134
 |||||
 99 TGATTCTTCTCAGCATGATGACCTGGCCACTGTCTGGGGAACCTCTCATCTCTT 158
 |||||
 135 caaatctatagacaaacctctgatacgccatgatagtttcttctctgacactgctgt 194
 |||||
 159 CATTAAGCATAGACTCCCGCTGCAACCCCATGTAATCTTCTCTCAGCAACATGTCTT 218
 |||||
 195 tctgagcatcatctgacacaaacacatcataccgaagatgctggagacatgctaacatc 254
 |||||
 219 TGATGAGAACTGCTTCTCCAC---CACCGTCCCAAGATGCTGGCAATCACAATCTCA 275
 |||||
 255 agaaatacattcatatgagctgcatgctccagctctcttcttctacatgctctct 314
 |||||
 276 GACTCAACACATCTCTCTCTGCACTGATGACGATATTTATCTGATGACCTGCG 335
 |||||
 315 gggagctgagatgctctcttcaacacacatgactgacagctatgctgacattgttt 374
 |||||
 336 TGACATGAGACATTTCTCTGCTGCTGATGAGCTATGACCGCTTGTCCCGTGTGCCG 395
 |||||
 375 cctcttctatcaagctatctatgaaacacacatgctgctgctgctgctgctgctgct 434
 |||||
 396 CCCCTTACATTAACACACAAAGATGACCCATGCTGCTGCTGCTGCTGCTGCTGCTG 455
 |||||
 435 catgctcatctgactgacacacatctctgctgctgctgctgctgctgctgctgctgct 494
 |||||
 456 ATGGGTGCTTCCCACTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
 |||||
 495 ctgtgggac 554
 |||||
 516 CTGTGCAAC 575
 |||||
 555 ctgtgacccctgtaagaatcaatgagtgatgctgctgctgctgctgctgctgctgctgct 614
 |||||
 576 CTGTTAC 635
 |||||
 615 agggagcattatcttactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 674
 |||||
 636 CACCCCATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 695
 |||||
 675 ccgcacagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 734
 |||||
 696 CCATCTCA--CAAAAGGAGATGAAGCCTTCTCCACCTGTGGCTCCACCTGTGGCTGAGGT 753
 |||||
 735 gacccctta 743
 |||||
 754 TCTCTCTTA 762

RESULT 9
 BH069789/c 698 bp DNA linear GSS 18-JUL-2001
 LOCUS
 DEFINITION RPI-24-331A10.TVB RPI-24 Mus musculus genomic clone
 ACCESSION BH069789
 VERSION BH069789.1 GI:14889386
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 698)
 Zhao,S., Nicrman,W., Malek,J., Shatsman,S., Akintel,B., Ievlins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 Other GSSs: RPI-24-331A10.TVB
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.tigr.org/tadb/bac-ends/mouse/bac_end_intro.html)
 page: http://www.tigr.org/tadb/bac-ends/mouse/bac_end_intro.html
 Plate: 331 row: A column: 10
 Seq primer: 77
 Class: BAC ends.

FEATURES

source location/Qualifiers
 1. 698
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-331A10"
 /clone_1ib="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTRABAC1. Site 1: BamHI. Site 2: BamHI.
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 230 a 150 c 169 g 149 t
 ORIGIN

Query Match 18.4%; Score 170; DB 12; Length 698;
 Best Local Similarity 52.7%; Pred. No. 6.2e-33;
 Matches 368; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

196 gtagacatcatctgcacaaagacatcatcaggaagctgaggacatgacatca 255
 |||||
 698 GTGACATCTGCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639
 |||||
 256 gaaatacattcatatgacagctgcatgctccagctctcttcttcaatgctctg 315
 |||||
 638 GGCACACATTTCTCAAGTTGCGTGTGAGTCAATCTTCTCTTACCTTCTTGTCT 579
 |||||
 316 ggaagctagatgctctcttcaacacacatgctgctatgacagctatgctgcttctt 375
 |||||
 578 TCTATTGACGTCTACCTCTTGGCAATCATGCTGACACCGCTATGAGCTGTGTGCCAA 519
 |||||
 376 ccttcttcaatgctatgataatgaacacacatgctgctgctgctgctgctgctgctgct 435
 |||||
 518 CCATTGCTTATGTCACATCATATGACAGAGCCCGGTAGGTGTGTAACGTGGGCA 459
 |||||
 436 atgctatgacagcaaatctctggtgacacagctcttcatatgaggtgacttctc 495
 |||||
 458 TATGTGGTGCTTTTTCAGAGTGCTTATTCGAGCTGACAGGCTTCACTTCTTCTTCT 399
 |||||
 496 tctgggccaacacacatgacacacttctctgtagatgaccccatgctgctgctgctgct 555
 |||||
 398 TGTGGAACATGAGATCAATTTATATGTAACCTCCGCCATTTGTTAAACCTGTA 339
 |||||
 556 tctgacccctgtaagaatcaatgagtgctgctgctgctgctgctgctgctgctgctgct 615
 |||||
 338 TGTGGGACACCTTACATTTCAAGAGGTGCTGATATTTGTTTGGCATTGTTTGTCTATGCT 279


```

Db      143 TCTCCTGCTGCGGCTTACACGCCCAAGTACTTCTCTGTCAGAGCTGCTGTTTGG 202
QY      200 acatcatctgcacaagaacatcaacgaagatctgtggagaccatgctaacaacagaa 259
Db      203 ACATATTTTCCCTTCTGAGATCCCTTAAGATGCTGCTATCTTTTAAATGAACAGCC 262
QY      260 ataccattcatcatcagagctcatgctccagctctctctgttcaatgctctcggag 319
Db      263 GAGCCATCTCTATGACAGGCTGATCCAGCTCTTCTTCAACATTTCTCGGGCTGCA 322
QY      320 ctgagatggtctctctccacacatgagcctatgagcagctatgtggcattgttccctc 379
Db      323 CTGAGTGTTCTCTGACAGGATGAGGCTTACAGCGCTTTGTGGCATTGTTCACCCG 382
QY      380 ttcatatcaagtaactattatgaaccacataatgtgtgagccttgcctgacagatgta 439
Db      383 TACGCTACACCAATATCAATGACGACACAGAGCGTATATCTATCCATGCTGAAGGACCTCAT 442
QY      440 ctatgtcagctcacaattcctgtgagacacagctcttcatgaggttgacttctgtg 499
Db      443 TCTTTGGCTGCAATTCAGGCACTTCTGACACCTCTCACTTCCATTTGCTTACTGTG 502
QY      500 ggccaacaacatgacacactctctctgtgagatacccccattgctgctgtcctgta 559
Db      503 GCCCAATGAGGTGACATATTTCTGATATCCAGTCACTGCTGAAGCTGGCTTGTG 562
QY      560 gccctgtaagaatcaatgagtgatgtgagtgtgctgtgatatccctggcacaatgag 619
Db      563 CAGATACCTTACCCCTGAGATGAGTGATGATCATGCTGAGGCTTCACTGAGCT 622
QY      620 acttattcttaccgtcactcctctgtgttattcaatgttgctatcttcctgata 675
Db      623 GTTCTCTTTTCACTCTACACTCTCTACAGCTGATATTGTGCTCATCTGACATC 678

```

RESULT 13

A2969227

LOCUS 642 bp DNA linear GSS 27-APR-2001
 DEFINITION 2M0241J24R Mouse 10kb plasmid U06C2M library Mus musculus genomic

ACCESSION

A2969227

VERSION

A2969227.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

```

/clone="U06C2M0241J24"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD42HV; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114[9]b1AF19072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 130 a 166 c 129 g 217 t
ORIGIN
Query Match 17.0%; Score 157.4; DB 12; Length 642;
Best Local Similarity 52.9%; Pred. No. 1e-29;
Matches 338; Conservative 0; Mismatches 301; Indels 0; Gaps 0;
QY 224 taacgaagatgctggagaccatgctaacatcagaataaccattcatatgagcgtgca 283
Db 1 TCCCCAGATGATGAGTGATCTTCTGTCAGAGAGAAATATCTCTATGAGAGCTGTG 60
QY 284 tgcacagctctctctgacatgctcctgagagctgagctgctcctcacacca 343
Db 61 TGACCCAGCTCTTTGATTCATTTCTTGTGGCTGACAGAGTCTCCCGGACGACGA 120
QY 344 tggcctatgacagctatgagcactgttccctcttcaatcagactatgagacc 403
Db 121 TGGCATATGATGATATATGCTATCTATGACCCGTTAAGTAGTATATGACGA 180
QY 404 accatatgctgtagcctgctcagcatgctatgctatgctgacacacattctcgg 463
Db 181 AGGCTCTGTCAGCTGCTGATGACGCTTATGCTGTCAGAGTGGTTTCTCACTCA 240
QY 464 tgcacagctcttcatcagagctgactctctgagggcacaacacacattgacactct 523
Db 241 TGCACAGGTTTGAACCTTCCACCTGCTTTGTGGTAAACATGACATCATTTATTT 300
QY 524 tctgtagataccacacatgctgctgtgctgtgagccctgtaagaatcaatgag 583
Db 301 TCTGTGACATATGATCTCTCTCTATCTTGTGATGATCTTCCCTCAATGAGACT 360
QY 584 tggtagatgctgatatattaccctggcacaatgaggaacttattcttaccctgac 643
Db 361 CTTTGTGTCATTTGGATCTCTCATATGCTGACATCTTCTGTCGCTCATCTTCT 420
QY 644 atggttttcatctgttgcattctcagcagcagtagaaggaaggaagagcct 703
Db 421 ACCTTACATATCTTCACCATCTGAGAGATCTGCTTCTCTGAGGAGGACCAAGCT 480
QY 704 tctcaacatgctcatctcatcagagtgagacatttactatctctctgtaactaca 763
Db 481 TTTCCACGCTGCTCCACCTGTCATATGATCTCTATATGAGAGTGTATCTTCA 540
QY 764 cctatatcgcctctctccagctatatacatattggaagagacagtgtagtgcacct 823
Db 541 CGTATGTGAGGCCATCTCATCTTACTCTTGAAGAAAGATGATGATGATCTCAAGT 600
QY 824 atactctgtgctccacacatcaaccgagtggttaca 862

```

FEATURES
 source
 1. 642
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

Db	601	ATAGTGTTCACACCCATGCTGAACTGTGAATTTATA	639
RESULT 14			
LOCUS	AO976178/c		
DEFINITION	AO976178 RPCI-23-333010.TV RPCI-23	Mus musculus	genomic clone RPCI-23-333010
ACCESSION	AO976178		
VERSION	AO976178		
KEYWORDS	GSS.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 600)		
	Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret		
	, B., Levins, M., Mognam, S., Tsegaye, G., Geer, K., Kroi, M., de Jong, P		
	and Fraser, C.M.		
	Mouse BAC End Sequences from Library RPCI-23		
	unpublished (1999)		
	Other GSS: RPCI-23-333010.TJ		
	Contact: Shaving Zhao		

FEATURES
source

Email: szhaoc@ligr.org
 Clones are derived from the mouse BAC library RPCT-23. For BAC
 clones available, please contact Pieter de Jong
 (pieter@edong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.ligr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 333 row: 0 column: 10
 Seq primer: T7
 class: BAC ends.
[View details](#)

Location/qualifiers
1. .600

```

1. 000
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-333010"
/clone_lib="RPCI-23"
/sex="Female"

```

BASE COUNT
ORIGIN

[illegible]

a	115 c	172 g
---	-------	-------

Query Match	16.9%;	Score 156.6;	DB 12;	Length 600;
Best Local Similarity	60.3%;	Pred. No. 1.6e-29;		
Matches 275; Conservative	0;	Mismatches 180;	Indels 1;	Gaps 1

QY	468	cacagctcttaacagatgattgactctctggtgcgaacaacacatgacacattcttcg	527
Db	585	CACGTGCTGAGAGACAGCGGTGTCATCTCTTGAGACCAAGGTCATCACCCGCTTCTTCG	526
QY	528	tggatataccccatctgcgtcgtttgctctgtagccctgaagaatcaatgagatgagt	587
Db	525	TGAGATTCCCCACACGCTTCTGCTCTCTGTAGCCCCACATAGGAAACAGATTATGAC	466
QY	588	gtatagtctgataattaccctgcgcataaggygaactatattcttaactgcattccctatgg	647
Db	465	TCTAATAGCAGATGTCCTTCATGGAGGAGCAATTTTGCTTACCTTACCTATCTCATGG	406

QY	648	ttttatcatgttgctattcttcggtalccgcacagtagaaggaaggaagggcccttc	707
Db	405	ctgcatactatggccagcactgctggccgacgcttctgctgaaggcagagagggcccttttc	346
QY	708	aactgctcatctatctatctccacagtggtgagcccttaactatctctcgttaatcaacta	767
Db	345	ttaacctcttcatcccaacctgcatcgtgctctctgcttactacatctgtgtctgctcta	286
QY	768	tatccgcctcgtcttccagctatatactatggaagagacaggtgtgtagctgtaactatac	827
Db	285	tgtacgcccctcatccacactataggcccgagaaagaaagaaacttaccctctgtgtgctac	226
QY	828	tcttgtagctcccaactaaacccgagtggtgtaagagcttccagatagagatgagcagc	887
Db	225	attctctcagcccaaccctggaacccctctaat-cttaccctttagagaaacaaagatgtcaacgt	167
QY	888	aggaattaggaagtgcttgtaattctctgaacacta	923
Db	166	ccgcattggcagacgctcttcttcctcttcttcacattta	131

[illegible]

REFERENCE
AUTHORS

JOURNAL, TITLE Creation of genome activation of gene expression Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athervs. Inc.

FEATURES

Email: scaln@ethersys.com
High quality sequence stop: 544

```

/cell_line="H1080"
/creation_of_genome_wide_protein_expression=
/notes="See 'Creation of Genome-wide Protein Expression'

```

/note="See 'Creati
Libraries using Ra

cell type indicated is HRI1080, since a random method was used, these sequence tags are not necessarily expressed in HRI1080 under normal circumstances."

BASE COUNT	ORIGIN	Score	DB	Length
Query Match	16.8%;	155.6;	10;	796;
Best Local Similarity	51.4%;	Pred. No 3.2e-29;		
Matches	384;	Conservative	0;	Mismatches 360; Indels 3; Gaps 1
4	aatcacagcgtgtaactgaatcattatctctgagccctacacaaaagcctgaactccag	63		
25	AAACCCACCTGCATGCTCTGAAATTCATCCCTCGTCTCTCTCTGCGCTTCCCCACCTCCAA	84		

THIS PAGE BLANK (USPTO)


```

|||||.....|
214 TTGCTCTCTCGGAGTGGCTATGCTTGTCTGTCATACCAAGATGCT 263
79 ugiYrHmetLeuThrSerLysnHrIleSerYrAlaGlyCysMetS 96
|.....|
264 GCAGAGCGCTTGTGAGTACGAGCCAGAGATCTTAGGAGGAGATGGCCA 313
96 ergHnLeuPheLeuPheTrpSerLeuGlyAlaGlyMetValLeuPhe 112
|.....|
314 CACGATGTTTTTCTTCACATTTTGGCAATGAGTGGCTGCTGCTATG 363
113 ThrHrMetAlaYrAspArgYrValAlaIleCysPheProLeuHnIstY 129
|.....|
364 GCAGCAGTGGCTTGTGACCGCTGATGGCATATGATGCTCCCTCCACTA 413
129 rSerThrIleMetLsnHnIstMetCysValAlaLeu..... 141
|.....|
414 TGCACCCGAAATGATGCTGAGTATGCGCCATTGGCAATTGTTGAT 463
142 .....LeuSerMetValMetAlaIleAlaValThrAsnSerTrpValHis 156
|.....|
464 GGGGAAATGGATGATAGTATGATGCGAGACAGCAATTTN..... 504
157 ThrAlaLeuIleMetArgLeuThrPheCysGlyProAsnThrIleAspH 173
|.....|
505 .....ATTNCTCCTGAACTCTGTGTGACCTTGTGAGATAGACCA 545
173 spHepheCysGluIleProProLeuLeuAlaLeuSerCysSerProValA 190
|.....|
546 CTTCTTCTGTGACCTTCCACCTCTCCCTGCACTTGTCTGTGATACAT 595
190 rGllleAsnGlyValMetValYrValAlaIleAspIleThrLeuAlaIleGly 206
|.....|
596 CCCAAAGCGAGCTGCATCTTGTGCGCAATTCCTCTGTAATCTAGT 645
207 AspPheIleLeuThrCysIleSerYrGlyPheIleIleValAlaIleLe 223
|.....|
646 CCATTTTGTGATCCTTATTTATGTCAGAAATTCGCTGCGAGTGT 695
223 uArgIleArgHnValGlyGlyAsnArgYsAlaPheSerHrCysSerS 240
|.....|
696 GGGAGAGCTTCCACCTGAGGGGCGCATAACTCTTCCACCTGTTCT 745
240 ergHnLeuThrValIleThrLeuYrTrpSerProValIleYrThrYr 256
|.....|
746 CACACCTACTTGTAGTACACTCTTTATGCGCTGTGCTCTTACCTAT 795
257 IleArgProAlaSerSerYrThrPheGluArgAspLysValAlaIleAl 273
|.....|
796 TTGAGGCGCAAGTCTAGCCACTCACAGAGATGACAACTCTTGGCCT 845
273 aleuYrThrLeuValIleThrProThrLeuAsnProMetValYrSerPhe 290
|.....|
846 CTTCTTACACAGATGATGATCCTGCTGAGAACCTATCATCTACAGTCTA 895
290 lnsAsnArgLysMetGlnAlaGlyIleArgLysValPheAlaPheLeuLys 306
|.....|
896 GGAACAGAGAAAGTCAAGCAGCACTGAGAAACTCTGCGCTGAAAAA 945

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:us-08-748-506-5

seq_documentation_block:

```

; Sequence 5, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL

```

```

COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,506
FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,751
FILING DATE: 09-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 74940
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5600
TELEFAX: 312-616-5700
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-748-506-5

```

```

alignment_scores:
Quality: 622.00 Length: 308
Ratio: 2.777 Gaps: 2
Percent Similarity: 72.727 Percent Identity: 41.883

```

alignment_block:

US-09-975-308-9 x US-08-748-506-5 ..

Align seg 1/1 to: US-08-748-506-5 from: 1 to: 966

```

2 AsnHisSerValIleThrGluPheIleIleLeuLeuThrLysLysPr 18
|.....|
31 AATGATTGACTGTCACAACTTT.....GCAATGGCAAGTCTCTC 71
18 ogLLeuGlnGlyIleIlePheLeuPhe.....LeuIleValT 32
|.....|
72 TGAGGTCCCTGAGAAATGCTTCCCTGTTCAACCTCATCTCTCATGT 121
32 yrLeuValAlaPheLeuGlyAsnMetLeuIleIleIleAlaLysIleYr 48
|.....|
122 TCTTGTATACACTACAGAAATCTCTCATAGTCTGCTATTTGTACC 171
49 SerAsnThrLeuHnIstHrProMetYrValPheLeuLeuThrLeuAlaVal 65
|.....|
172 ACTGCATCTTACACACCCCACTGATCTTCTTGTGGCAACTGTGTCT 221
65 lValAspIleIleCysThrThrSerIleIleProLysMetLeuLysHn 82
|.....|
222 CCTGAGATTTGGCTATACTTCTCTGTCATACCCAAAGTGTGAGAGCC 271
82 etLeuThrSerLysnHnIstMetYrAlaGlyCysMetSerLysLeu 98
|.....|
272 TTGTGAGTGGAGCCAGAGATCTCTCGGAGGAGATGGCCACACAGAT 321
99 PheLeuPheThrTrpSerLeuGlyAlaGlyMetValLeuPheThrThme 115
|.....|
322 TTTTTCGCAATTTTGTGTAATGATGCTGCTATTTGGCGCAT 371
115 talatYrAspArgYrValAlaIleCysPheProLeuHnIstYrSerThr 132
|.....|
372 GGCCTTGAACCGCTGCATGCGCATATGCTCCCACTCCACTATGCAACC 421
132 leMetLsnHnIstMetCysValAlaLeuLeuSerMetValMetAlaIle 148

```

```

422 GAATGAGTCGGNAGGTATGTGCGCCATTGCGCAATTTGTTCAATGGGGAATG    ::::      :::
149 AIAValThraNSerTTPVAlHIShtAlaleuilemetargleuthrph 165
   ::: ||| ::::: ::::: ||::: |
472 GGATGCATAGTAAGACTGCTGGACAAACCATAATTTATTTCTCCTTGAACTT 521
165 eCySLyPProAnThrILeASPhISpHePCysGLUluIleProProLeuL 182
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
522 CTGTGGACCCCTGTGAATAGACACACTTCTTGTGTGACCTTCCACTCTCC 571
182 euAlAleuSerCySSerProValArGIleASngluValmeValTyrrVal 198
572 TGGCACTTGCCTGTGGAGATCATGCCCAAACAGAGCGCCATCTTGTGTG 621
199 AlAspIleThLeuAlIlEGlyAspheIleLeuThrcySleSerty 215
   ::::: ::::: ::::: ||::: ||::: |||||
622 GAGACAGTCCTCTGCATATCTAGCCCATTTTTGCTGATCATTTATTTCTTA 671
215 rElYpHeIIleValAlAlAleuArgIleArgThrValGluglyLSA 233
   | |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
672 TGTCAAAATTTCTCATTTGCAAGTGCNTGATGCTTCACTCACGTAGAGGGGCC 721
233 rGLYSAlAPheserThrcySSerSerHisLeuThrValValThLeuTyrr 248
   ::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
722 ATMAAGCTCTTCCACCTGTTCGCTCCACCTACCTGTAAGTACACACTTTT 771
249 TyrsSerProValIlleTyrrThrTyrlleArgProAlasSerTyrrThrh 265
   |||||::: ::::: |||||::: |||||::: |||||::: |||||::: |||||:::
772 TATGGCTCAGCATGTATTACTATTGTAAGGCCCAAGTCTAGCCACTAAC 821
265 eGUArGspLySVAlValAlAlAleuTyrrThrleuValThrProThrl 282
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
822 AGGAATGACAAATTCCTGGCCCTCTTCTACACAGTAgTGCATCCATGTC 871
282 euASPrOmeValTyrrSerPheGlnASnAsglUmEtGlnAlaglyLe 298
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
872 TGAACCTATCATCTATATGTTTAAAGAACAAGATCAAGCGACACTG 921
299 ArgLySVAlPheAlAlPheLeuLys 306
   |||||::: ::::: |||||
922 AGAAGACTGTGGGCTGAAAAA 945

seq_name: /cgr2_6/prodata/1/ina/6A_COMB.seq:US-08-748-506-7

seq_documentation_block:
? Sequence 7, Application US/08748506
? Patent No. 6159707
GENERAL INFORMATION:
? APPLICANT: Romnett et al.
? TITLE OF INVENTION: NOVEL SPERM RECEPTORS
? NUMBER OF SEQUENCES: 31
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Leydig, Volt & Mayer, Ltd.
? STREET: Two Prudential Plaza, Suite 4900
? CITY: Chicago
? STATE: IL
? COUNTRY: US
? ZIP: 60601-6780
COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/748,506
? FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/033,751
? FILING DATE: 09-NOV-1995
CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
```

```

: REFERENCE/DOCKET NUMBER: 74940
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5600
: TELEFAX: 312-616-5700
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-748-506-7

alignment_scores:
      Quality: 620.50      Length: 296
      Ratio: 2.846      Gaps: 3
Percent Similarity: 73.649      Percent Identity: 44.257

alignment_block:
US-09-975-308-9 x US-08-748-506-7  ..

Align seg 1/1 to: US-08-748-506-7 from: 1 to: 966

16 lysLysPROGluLeuGlnGlyIleIlePheLeuPhe.....Le 29
   ||| |||::: ||| |||::: ||| |||
64 AAGTCTCTGAGGTCCCGGAGAAATGCTCTCTCCGTACACCCTCAACCT 113
29 ulleValIyLeuValAlaIlePheLeuGlyAsnMetLeuIleIleIleAl 46
   : ::::: |||::: |||::: |||::: |||::: |||::: |||:::
114 TCTCATGTTCTTAGTACACTACACAGAAATACCTCTCATAGCCTTGCTA 163
46 ysIleTySerAsnThrIleuAsnThrIleuPheLeuLeuThr 62
   ||| ||| |||::: |||::: |||::: |||::: |||::: |||:::
164 TTGTGACACAGTCACCTCTACACACACCCCACTGTAATCTTCTTGCCAC 213
63 leuAlaValAlaAspIleIleCysThrThrSerIleIlePheProLysMet 79
   |||::: |||::: ||| ||| |||::: |||::: |||::: |||::: |||
214 TTGTCTCTGCGAATGGATGGCTATACCTGCTCTGTCAACCCCAAGATGCT 263
79 uGlyThrMetLeuThrSerGluAsnThrIleSerTyAlaGlyCysMet 96
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
264 GCAGAGCCCTGTGAGATGAGGCCGACGAGATCTCTGGAGGGTGTGCCT 313
96 ergIleuPheLeuPheThrTrpSerIleuGlyAlaGluMetValLeuPhe 112
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
314 CACAATGTTCTTCTTCATATTCCTTGGTATACAGTGGCTGCTCATATG 363
113 ThrThrMetAlaTyAspArgTyValAlaIleCysPheProLeuHisTy 129
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
364 GCAGCCATGGCCCTTGGACCCCTATATGCTATATGTTCCCACTCCACTA 413
129 rSerThrIleMetAsnHisMetCys.....ValAlaLeuLeuSer 144
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
414 TGCACACCCGAATGAGATCGCTGGGGATGCTGCTATTTGGCAATTTGCAT 463
144 eValMetAlaIleAlaValThrAsnSerTrpValHisThrAlaLeuIle 160
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
464 GGGTATGGAGCATACTA.....GGCTGGACACACCAATTTTAAT 507
161 MetArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCysG 177
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
508 TTCTCCTTGAACTCTGTGAGACCGCTGTGAGATACACCACTCTCTGTGA 557
177 uIleProLeuLeuAlaLeuSerCysSerProValArgIleAsnGluV 194
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
558 CCTTCCACCTCTCCGAGCACTGCTGTGGATGATACCCCAAAATAGAG 607
194 AlMetValTyValAlaAspIleThrIleuAlaIleGlyAspPheIleLe 210
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
608 CTGCACATCTTTTGTGCACACAGTGTCTGTGCATATTAGTCCATTTTAC 657
211 ThrCysIleSerTyGlyPheIleIleValAlaIleLeuArgIleArgTh 227

```

```

|||||
658 ATCATTTCTCATGTCAGAAATTCGTGTCAGTGTGTCATGTCCTTC 707
227 rValGluglyLysArglySalapheserThrcysSerSerHisLeuThrv 244
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 ACCGAGGGGGCCATTAAGCTCTGTACTGTGTCTGTCACCTACTG 757
244 aValThrLeuThrvThrcysProValIleThrvThrvIleArgProAla 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TAGCACACTCTTCTATGCGTCACACATCGCACCTATTGAGGTCCAG 807
261 SerSerThrvThrvPhgGluArgAspLysValAlaIleAlaLeuThrv 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
808 TCTAGCCACTCACCCAGAGTGGACAACTCTGGCCCTCTCTATACATC 857
277 uValThrProThrLeuAsnPrometValIleThrvSerPhgGlnAsnArgLum 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
858 AGTACATCCATGCTGAATCCCATCATCTACACGCTTAAGGACAGGAG 907
294 etGlnAlaGlyIleArgLysValPheAlaPheLeuLys 306
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
908 TAAAGGCTGCATGAGAAAGAACTCTGGGCTGAGAGAAA 945

```

seq_name: /cgn2_6/prodata/1/ina/6A_COMB.seq:US-08-748-506-8

```

seq_documentation_block:
; Sequence 8, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-748-506-8

```

```

alignment_scores:
  Quality: 614.50      Length: 296
  Ratio: 2.845        Gaps: 3
Percent Similarity: 72.973  Percent Identity: 43.581
alignment_block:

```

US-09-975-308-9 x US-08-748-506-8 ..
Align seg 1/1 to: US-08-748-506-8 from: 1 to: 966

```

16 LysLysProGlnLeuGlnGlyIleIlePheLeuPhePhe.....Le 29
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
64 AAGTCTCTGAGAGTCCCTGGAGAAATGCTCTCTGTTGACCCCTCATCT 113
29 uIleValIleThrvValAlaPheLeuGlyAsnMetLeuIleIleAlaL 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 TCTCATGTTCTTATGATCTACAGAGAAATGCTCTCATAGCCCTGGCA 163
46 ySleIleThrvSerAsnThrLeuHisThrPrometThrvValPheLeuThrv 62
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
164 TTGTGACAGATCCATGCTTACACACCCCATGACTTCTTCTTGGCCAC 213
63 LeuAlaValAlaAspIleIleCysThrvThrvSerIleIleProLysMet 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 TTGTCCTCTCTGGAGATGGCTATGCTTGTGCTCTGTGTCATACCAATG 263
79 uGlyThrMetLeuThrvSerGlnAsnThrIleSerThrvAlaGlyCysMet 96
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 GCAGAGCTCTGTGAGTGGAGGCCGAGAGATCTTTCAGTGGGATGGCCA 313
96 eArgLLeuPheLeuPheThrvThrvThrvSerIleGlyAlaGluMetVal 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 CACAGATGTTTCTTCTCATATCTTGTGTAATACAGAGTGGCTGCTATTG 363
113 ThrThrvMetAlaThrvAspArgThrvValAlaIleCysPheProLeuHis 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 GCAGCCATGGCTTGTGACCCCTATATGCTATATGTTCCCACTCCACAT 413
129 rSerThrvIleMetAsnHisHisMetCys.....ValAlaLeuLeuSer 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 TGCACACCAATGAGTGTGAGTATGTGCCACCTGGCAATGTTTCAT 463
144 eValIleMetAlaIleAlaValThrAsnSerThrvValHisThrAlaLeu 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 GGTGATGGATGCATGTA.....GGTCTGGACAGACCAATTTAT 507
161 MetArgLeuThrvPheCysGlyProAsnThrIleAspHisPhePheCysG 177
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
508 TTTCCTTGAACCTCTGTGACCTGTGAGATACCACTCTCTCTGTGA 557
177 uIleProProLeuLeuAlaLeuSerCysSerProValArgIleAsnGlu 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
558 TCTTCCACCTCTCTGACACTTGTGTGTGATACATCCCAATGTAGG 607
194 alMetValThrvValAlaAspIleThrvLeuAlaIleGlyAspPheIle 210
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
608 CTGCATCTTGTGTAGTGTCTGCTGTCGATATCTAGCCCTTTTGTCTG 657
211 ThrCysIleSerThrvGlyPheIleIleValAlaIleLeuArgIleArg 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
658 ATCATTTATCTTATGTCAGAAATCTCGTTCAGAGGCTGGATGCTTC 707
227 rValGluglyLysArglySalapheserThrcysSerSerHisLeuThrv 244
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 ACCGAGGGGGCCACAAAGCCCTTCAACCTGTCTCCACCACTACTG 757
244 aValThrLeuThrvThrcysProValIleThrvThrvIleArgProAla 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 TAGTACACTCTTATGCTGCTCAGATCTGTACCTATTGAGGCTAG 807
261 SerSerThrvThrvPhgGluArgAspLysValAlaIleAlaLeuThrv 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
808 TCTAGCCACTCACCCAGAGTGGACAACTCTGGCCCTCTCTACACAGC 857
277 uValThrProThrLeuAsnPrometValIleThrvSerPhgGlnAsnArgLum 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
858 AGTACATCCATGTTGAACCTATCATCTATATGTTAAGGACAGGATG 907
294 etGlnAlaGlyIleArgLysValPheAlaPheLeuLys 306

```

908 TCAAGCAGCAGCTGAGAGATTCCTGCGCCCTGAGAAAA 945

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-827-291A-1

seq_documentation_block:

; Sequence 1, Application US/08827291A

; Patent No. 5874243

; GENERAL INFORMATION:

; APPLICANT: Macina, Roberto

; APPLICANT: Sathe, Ganesh

; TITLE OF INVENTION: NOVEL OLRCCL5 RECEPTOR

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/827,291A

; FILING DATE: 28-Mar-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATORNEY/AGENT INFORMATION:

; NAME: King, William T

; REGISTRATION NUMBER: 30,954

; REFERENCE/DOCKET NUMBER: GP50001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-5015

; TELEFAX: 610-270-5090

; TELETYPE:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1290 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-827-291A-1

; alignment_scores:

; Quality: 598.00 Length: 302

; Ratio: 2.682 Gaps: 2

; Percent Similarity: 73.841 Percent Identity: 36.755

; alignment_block:

; US-09-975-308-9 x US-08-827-291A-1 ..

; Align seg 1/1 to: US-08-827-291A-1 from: 1 to: 1290

; 2 AsnHsSerValValThrgluPheIleIleuGlyLeuThrIleValVal 18

; 308 AATTGCACTTCACTCGACTTCATCTCTGGAAATCTTCAATGACAG 357

; 18 ogIuLeuGlnGlyIleIlePheLeuPhePheLeuIleValIleVal 35

; 358 CCCACCCACACCTCTCTCTTCTTGTGCTGGCCATCTTTTCAGTGG 407

; 35 IarPheLeuGlyAsnMetLeuIleIleIleAlaIleValIleValSerAsnThr 51

; 408 CCTTACAGGAGAACTGTCTCATGTCTCTCATCTTCACTGACACACAG 457

; 52 LeuHsIthrGrometIleValIlePheLeuLeuThrLeuAlaValAlaAsp.. 67

458 CTCACACACCCCATGATGATCTCTCTGACGCACTGTGCTCCATGAGACT 507

68 ...IleIleCysThrThrSerIleIleProIleGlyMetLeuGlyThrMetI 83

508 CATGCTCATCTGACACAC...GTACCCAGATGGCTTCAATACAC 551

83 eutHrSerGluAsnThrIleSerIleIleGlyMetLeuGlyThrMetI 99

552 TGCTGGCAGCAGTCAATTCATGCTGCTGTGTGACACAAATTTTC 601

100 LeuPheThrIlePheSerLeuGlyAlaIleMetValLeuPheThrMetI 116

602 TTCTATACATCAGCTGCTGCTGATGCTTTTGTGCTGTATGCG 651

116 aTyAspArgIleValAlaIleCysPheProLeuHsIleThrIleM 133

652 TTATGACCGGTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701

133 etAsnHsIleMetCysValAlaLeuLeuSerMetValMetAlaIleAla 149

702 TGAGCCCTAAATTTGTGACTTATGACTGCTTTCTGAGTCTGGGC 751

150 ValIleAsnSerIleValIleThrAlaLeuIleMetArgLeuThrPheC 166

752 TCTACAGATGATCATCTTATGCTGATGACCAATTTCTCTCTGCTG 801

166 sGlyProAsnThrIleAspHisPhePheCysGluIleProProLeuAla 183

802 TGGCTCTGGGAAATGACCCATCTTCTGATGATGATGATGATGATG 851

183 IaleuSerCysSerProValArgIleAsnGluValMetValIleValAla 199

852 TCTCTCATGATGATGACATCATATTTGAAAGTTATTTTCATTTGC 901

200 AspIleThrLeuAlaIleGlyAspPheIleuThrCysIleSerIleVal 216

902 TCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951

216 yPheIleIleValAlaIleLeuArgIleArgIleValGluGlyAsnArg 233

952 TGGATTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1001

233 yAlaPheSerThrCysSerSerHisLeuThrValIleThrLeuThrTyr 249

1002 AAGCTTACAGACCTGTTCTCTCATGCTGCTGCTGCTGCTGCTGCT 1051

250 SerProValIleThrIleArgProIleAsnSerIleThrPheG 266

1052 GGACGAGCTTGTTCATGATCATGACACCCACATCATGCTGCTGCCAAC 1101

266 uArgAspIleValAlaIleLeuThrIleValIleThrIleValIleThr 283

1102 GAGGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1151

283 snProMetValIleThrPheGlnAsnArgGluMetGlnAlaGlyIleArg 299

1152 ATCCCTCATCTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1201

300 LysVal 301

1202 AAGATC 1207

seq_name: /cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-085-371-5

seq_documentation_block:

; Sequence 5, Application US/09085371

; Patent No. 6218358

; GENERAL INFORMATION:

; APPLICANT: Firestein, Stuart

; APPLICANT: Zhao, Haiding

; TITLE OF INVENTION: Functional Expression of, and Assay for, Functional Cellular

; TITLE OF INVENTION: Vivo

```

197   tyValAlaaspIlethrLeuAlaIleGlyAspPheIleThrCysII 213
      ::::| | | | | ::::| | | | | ::::| | | | |
613   TTtTtCCtGGcCAATTTTAATCTGTGGGAACCGCTCTGTGCACtGGGGC 662
      ::::| | | | | ::::| | | | | ::::| | | | |
213   eSeTYrGLyPhelIeIleValAlaIleLeuArgIleaGrHValGluc 230
      ||||| | | | | | | | | | | | | | | | | |
663   ATCTCAATGAGCCATCACAGAGTCGTCTGATGGCAACCCCTCAGCTGCTG 712
      ::::| | | | | ::::| | | | | ::::| | | | |
230   IyLysArGLysAlaPheSerThrCyseSerHisLeuThrValAlaThr 246
      ::::| | | | | ::::| | | | | ::::| | | | |
713   GCGCGCAATAAAGCCTTTCAACCTGTGCCCTCCACCTCATCTGTGTGATC 762
      ::::| | | | | ::::| | | | | ::::| | | | |
247   LeuTyTrYrSerProvaIIIEtyrThrTYrIleArgProAlaSerSetry 263
      ::::| | | | | ::::| | | | | ::::| | | | |
763   ATCTTCTATGcAGCCACTATTTCATCTATGACCAGGCGCTTAAGSACAATCC 812
      ::::| | | | | ::::| | | | | ::::| | | | |
263   rThPheGIuArgspLySValAlaAlaLeuTYrThrLeuValThrP 280
      ::::| | | | | ::::| | | | | ::::| | | | |
813   AGCTTTTACACCAACAAGCTGGCTGTGTACTGACTCAGCTGTCATTGTATC 862
      ::::| | | | | ::::| | | | | ::::| | | | |
280   rOrHLeuAsnProMeValTYrSerPheGIuAsnArg 292
      || | | | | | | | | | | | | | | | |
863   CFTTGTCATCCATCATCTACTACGCTTGTGGCAACCAA 900

```


MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 116..1003
 US-08-467-948A-1

alignment_scores:
 Quality: 577.50 Length: 297
 Ratio: 2.831 Gaps: 5
 Percent Similarity: 68.687 Percent Identity: 42.424

alignment_block:
 US-09-975-308-9 x US-08-467-948A-1 ..

Align seg 1/1 to: US-08-467-948A-1 from: 1 to: 1713

```

2 AsnHisSerValAlaThrGluPheIleIleLeuGlyLeuThrLysLysPr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 AATCAGACAGATGGTCACAGAGTTCTCTCTACTGAGATTTCCTCGGGCCC 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18 OGLeuGlyGlyIleIlePheLeuPhePheLeuIleValIleuVala 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 AAGGATTCAGATGCTCTCTTGGGCTCTCTCCCTGTCTATGTCTCA 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleIleYrSerAsnThr 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
237 CCGCGCTGGGGAAGGAGACCATCTGGGCTCATCTGACTGACTCCAGA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 LeuHisThrProMetIleValIlePheLeuLeuThrLeuAlaValIaAspI 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
287 CTCACACCCCCCATGTACTCTTCTCTCTACACCTGGCCGCGTCAACAT 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrS 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
337 CGCCTATGCTGCACACAGAGTCCCGCATGCTGTAACCTCCTCGATC 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85 ergLuanThrIleSerIleAlaGlyGlyMetSerGlnLeuPheLeu 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
387 CAGCCAGCCCATCTCTGCTGGTGGTGCATGACATGACTATCTCTCTT 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaYrAs 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
437 TTGAGTTTGCACATCTGATGATGCTCTGCTGCTGATGCTCTACGA 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
118 ParGlyValAlaIleCysPheProLeuHisIleYrSerThrIleMetAsnH 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
487 TCGTAGCTGGCCATCTGCCACCTCTCGATATTTCATCATCATGACT 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
135 IsthMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 GGAAGCTCTCATCATCTCTG.....GGCATACT 565
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 AsnSerTrp.....ValHisThrAlaLeu 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
566 ...TCCGGACATGCTGCTCCCTCTGCTATGCTGATGATGAGCCCAT 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 eMetArgLeuThrPheCysGlyProAsnThrIleAspHisPhePheCysG 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 CCAAGACTCCCTTTGCTGGCTGCTGAATCAACACCTCTCTCTGTG 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
177 IuIleProProLeuLeuAlaLeuSerCysSerProValArgIleAsnGlu 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 AAATCTGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
194 ValMetValIleValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
713 GTGGTCATCTTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 uThrCysIleSerIleGlyPheIleIleValAlaIleLeuArgIleArgT 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
763 GGTGCTGCTCTCTACTACACATCTCTGCTGCTGCTGCTGCTGCTGCT 812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

227 hrValGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThr 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
813 CTGGGAGAGGGCGGAGAAAGCCTTCTCCACCTCTCTCCACACTGCG 862
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
244 ValValThrLeuIleYrSerProValIleIleYrThrIleArgProAl 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
863 GTAGTGGAGACTCTTCTTGGGAGCGGCATGCTGATGATGAGCCCTAA 912
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 aSerSerIleThrPheGluArgAspLysValVal...AlaAlaLeuYrT 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
913 GTCCGCCATCTGAGAGAGCAGAAAGTCTTTTCTTATTTTACAGT 962
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 hrLeuValThrPro...ThrLeuAsnProMetValIleYrSer 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
963 TCTTTCAACCCCATGCTTAACCCCTGATTTACAAC 1001
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/plodata/1/lna/6A.COMB.seq:US-08-467-947A-1

seq_documentation_block:

Sequence 1, Application US/08467947A

Patent No. 6090575

GENERAL INFORMATION:

APPLICANT: LI, YI

APPLICANT: CAO, LIANG

APPLICANT: NI, JIAN

APPLICANT: GENTZ, REINER

APPLICANT: BUTT, CAROL J.

APPLICANT: SUTTON III, GRANGER G.

APPLICANT: ROSEN, CRAIG A.

TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein

TITLE OF INVENTION: Coupled Receptor GPR1

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVE., NW, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,947A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

Prior APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04079

FILING DATE: 30-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.1140002/EKS/KLM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1713 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 116..1003

US-08-467-947A-1

alignment_scores:
 Quality: 577.50 Length: 297

Ratio: 2.831 Gaps: 5
Percent Similarity: 68.687 Percent Identity: 42.424

alignment block:

US-09-975-308-9 x US-08-467-947A-1 ..

Align seg 1/1 to: US-08-467-947A-1 from: 1 to: 1713

```

2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThyLysPr 18
137 AATGACACATGTCTACAGAGTTCTCTCTACTGGGATTTCTCTGGCCC 186
18 OgluLeuGlnGlyIleIlePheLeuPhePheLeuIleValThyLeuVala 35
187 AAGGATGAGATGCTCTCTCTGGGCTCTCTCTCTCTCTCTCTCTCTCA 236
35 lApheLeuGlyAsnMetLeuIleIleIleAlaLysIleTySerAsnThr 51
237 CCTCTGGGAGATGGACCATCTGGGGCTCATCTGACCTGACCTCAGA 286
52 LeuHisThrProMetTyValPheLeuLeuThrLeuAlaValAsp11 68
287 CTCCACACCCCATGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
68 eIleCyThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrs 85
337 GCCTATGCTGACACACAGTCCCATGATGCTGTAACCTCTGATC 386
85 eGluAsnThrIleSerTyValGlyMetSerGlnLeuPheLeuPhe 101
387 CAGCCACACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 436
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrMetAlaTyAs 118
437 TTGAGTTTGCACATGTAATGCTCTCTCTCTCTCTCTCTCTCTCTCA 486
118 PaGlyTyValAlaIleCyPheProLeuHisTySerThrIleMetAsnH 135
487 TGGTACTGGGCTGATGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCA 536
135 lSHisMetCyValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
537 GGAAGTGTGATCATCTCTG.....GGCATCT 565
152 AsnSerTrp.....ValHisThrAlaLeu1 160
566 ..TCCTGACATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 612
160 eMetArgLeuThrPheCyGlyProAsnThrIleAspHisPheCySG 177
613 CTAAGACTGCCCTTTGTGGGCTCTGTAATCAACCATCTCTCTCTCTG 662
177 lIuIleProLeuLeuAlaLeuSerCySerProValArgIleAsnGlu 193
663 AAATCCTGTCTGTCTGAGCTGCTGTCTGTCTGTCTGTCTGTCTGTCT 712
194 ValMetValTyValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
713 GTGTCATCTTGAAGCTGATCTCTGTGTGTGTGTGTGTGTGTGTGT 762
210 uThrCysIleSerTyrglyPheIleIleValAlaIleLeuArgIleArg 227
763 GGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
227 hrValGluGlyLysArgLysAlaPheSerThrCysSerHisLeuThr 243
813 CTGGGAGAGGCGGCAAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCT 862
244 ValValThrLeuTyTySerProValIleTyThrTyIleArgProAl 260
863 GTAGTGGAGACTCTCTTGTGGAGGCGCATGCTATGATGAGGCCCTAA 912
260 aSerSerTyThrPheGluArgAspLysValVal..AlaAlaLeuTyTr 276

```

```

913 GTCCCGCATCTGAGGAGCAGAGAGTCTCTTTCTTTTATCTACT 962
276 hrLeuValThrPro.ThrLeuAsnProMetValTySer 288
963 TCTTTTCAACCCCGATGCTTAACCCCTGATTTTCAAC 1001

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-748-506-9

```

seq_documentation_block:
; Sequence 9, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-748-506-9

```

alignment_scores:
Quality: 524.00 Length: 305
Ratio: 2.449 Gaps: 0
Percent Similarity: 70.164 Percent Identity: 33.115

alignment_block:
US-09-975-308-9 x US-08-748-506-9 ..

Align seg 1/1 to: US-08-748-506-9 from: 1 to: 984

```

2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThyLysPr 18
137 AATGACACATGTCTACAGAGTTCTCTCTACTGGGATTTCTCTGGCCC 186
18 OgluLeuGlnGlyIleIlePheLeuPhePheLeuIleValThyLeuVala 35
187 AAGGATGAGATGCTCTCTCTGGGCTCTCTCTCTCTCTCTCTCTCTCA 236
35 lApheLeuGlyAsnMetLeuIleIleIleAlaLysIleTySerAsnThr 51
237 CCTCTGGGAGATGGACCATCTGGGGCTCATCTGACCTGACCTCAGA 286
52 LeuHisThrProMetTyValPheLeuLeuThrLeuAlaValAsp11 68
287 CTCCACACCCCATGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
68 eIleCyThrThrSerIleIleProLysMetLeuGlyThrMetLeuThrs 85
337 GCCTATGCTGACACACAGTCCCATGATGCTGTAACCTCTGATC 386
85 eGluAsnThrIleSerTyValGlyMetSerGlnLeuPheLeuPhe 101
387 CAGCCACACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 436
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrMetAlaTyAs 118
437 TTGAGTTTGCACATGTAATGCTCTCTCTCTCTCTCTCTCTCTCTCA 486
118 PaGlyTyValAlaIleCyPheProLeuHisTySerThrIleMetAsnH 135
487 TGGTACTGGGCTGATGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCA 536
135 lSHisMetCyValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
537 GGAAGTGTGATCATCTCTG.....GGCATCT 565
152 AsnSerTrp.....ValHisThrAlaLeu1 160
566 ..TCCTGACATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 612
160 eMetArgLeuThrPheCyGlyProAsnThrIleAspHisPheCySG 177
613 CTAAGACTGCCCTTTGTGGGCTCTGTAATCAACCATCTCTCTCTCTG 662
177 lIuIleProLeuLeuAlaLeuSerCySerProValArgIleAsnGlu 193
663 AAATCCTGTCTGTCTGAGCTGCTGTCTGTCTGTCTGTCTGTCTGTCT 712
194 ValMetValTyValAlaAspIleThrLeuAlaIleGlyAspPheIle 210
713 GTGTCATCTTGAAGCTGATCTCTGTGTGTGTGTGTGTGTGTGTGT 762
210 uThrCysIleSerTyrglyPheIleIleValAlaIleLeuArgIleArg 227
763 GGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
227 hrValGluGlyLysArgLysAlaPheSerThrCysSerHisLeuThr 243
813 CTGGGAGAGGCGGCAAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCT 862
244 ValValThrLeuTyTySerProValIleTyThrTyIleArgProAl 260
863 GTAGTGGAGACTCTCTTGTGGAGGCGCATGCTATGATGAGGCCCTAA 912
131 CCTCATGGGCAACATGCTCTATTAATTTACATCTCTGTGTGTGTGTGT 180

```

[illegible]

1 APPLICANT: Yue, Henry
2 TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED
3 WITH IMMUNE RESPONSE
4 NUMBER OF SEQUENCES: 9
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: Incyte Pharmaceuticals, Inc.
7 STREET: 3174 Porter Drive
8 CITY: Palo Alto
9 STATE: CA
10 COUNTRY: USA
11 ZIP: 94304
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Diskette
14 COMPUTER: IBM Compatible
15 OPERATING SYSTEM: DOS
16 SOFTWARE: FASTSEQ for Windows Version 2.0
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/988, 876
19 FILING DATE: Herewith
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER:
23 ATTORNEY/AGENT INFORMATION:
24 FILING DATE:
25 NAME: Billings, Lucy J.
26 REGISTRATION NUMBER: 36,749
27 REFERENCE/DOCKET NUMBER: PF-0441 US
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 650-855-0555
30 TELEFAX: 650-845-4166
31 TELEX:
32 INFORMATION FOR SEQ ID NO: 2:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 1828 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 LIBRARY: PROSNOT01
40 CLONE: 364702
41 US-08-988-876-2

alignment_scores:	
Quality:	399.50
Ratio:	1.958
Percent Similarity:	64.353
	Percent Identity: 37.855

```
alignment_block:
US-09-975-308-9 x US-08-988-876-2 ..
```

Align seg 1/1 to: US-08-988-876-2 from: 1 to: 1828

[illegible]

```

85  ergluasnThrIleSerTyrAlaGlyCysMetSerGlnLeuPhe 101
      ::::::::::::::::::::
1043 ACACACAGATCTCTATGACAGCGCCGACTGAGATGCTCTCTT 1092
102  ThrTrpSerLeuGly...AlaGlnMet.ValLeuPheThrThrMetIat 117
      ::::::::::::::::::::
1093 GCCATTTTGGAGGCATGAGAGACATGCTCTGAGTGATGGCCCT 1142
117  YRPARATGTYrValAlaIleCysPheProLeuHisTyrSerThrIleMet 133
      ::::::::::::::::::::
1143 ATGACTGGTTGTAGCCATCTGTACCCGCTATATCATTC. ACCATCATG 1191
134  AsnHisHisMetCys.....ValAlaLeuLeuSerMetValMetAlaIle 148
      ::::::::::::::::::::
1192 AACCGGTGTTCTGTGCTTCTAGTTTGTTGTTCTTTTCTTTT...CT 1238
148  eAlValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThr. 164
      ::::::::::::::::::::
1239 CAGTCTTTTAACTCCACGCTGCACACACTTGATGGCTTACAGTGACCT 1288
165  .....PheCysGlyProAsnThrIle 171
      ::::::::::::::::::::
1289 GCTTCACAGATGTGGAAATTCCTAATTTCTCTGTGACCCCTTCACACTC 1338
172  AspHisPhePheCysGlnIleProProLeuLeuAlaLeuSerCysSerPr 188
      ::::::::::::::::::::
1339 TCCCATCTTGCACTG.....TGAGAAC 1361
188  oValArgIleAsnGlnValMetValTyrValAlaAspIle.....ThrL 203
      ::::::::::::::::::::
1362 CTTCACCAATTAACATATCATATGATTCCTGCTGCATATGTTGGTTTC 1411
203  euAlaIle...GlyAspPheIleLeuThrCysIleSerTyrGlyPheIle 218
      ::::::::::::::::::::
1412 TTCCCATCTTCAGGACCTTCTCTTACTGTA.....AAAT 1449
219  IleValAlaIleLeuArgIleArgThrValAlaGlnGlyLysArgLysAlaPh 235
      ::::::::::::::::::::
1450 CTTTCTCTCATTTCTGAGGTTTCATCATCAGGTGGGAAGTATAAACCT. TT 1498
235  eSerThrCysSerSerHisLeuThrValAlaThrLeuTyrTyrSerProV 252
      ::::::::::::::::::::
1499 CTCACCTGTGGGTCTCACCTGTGAGTTGTT...TGCTGGAGGTAACCTC 1545
252  aIleLeuThrTyrIleArgProAlaSerSerTyrThrPheGlnArgAsp 268
      ::::::::::::::::::::
1546 GGTTCCAGATGTGTCATCTTCCC.....GAGAAAGAG 1577
269  LysValAlaIleAlaLeuTyrThrLeuValThrProThrLeuAsnProM 285
      ::::::::::::::::::::
1578 TGCAGTGGCGCTCAGTGATGACAGGTGGTGCACCCCATGCTGAACCCCT 1627
285  eValIleTyrSerPheGlnAsnArgGlnMetGlnAlaGlyIleArgLys 300
      ::::::::::::::::::::
1628 TCATGTACAGCCCTGAGAAACAGGATATGAAAGTCTCGCGGCGG 1674
seq_name: /cgn2_6/ptodata/1/fna/6B_COMB.seq:US-09-439-313-526
seq_documentation_block:
: Sequence 526, Application US/09439313
: Patent No. 6329505
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan Louise
: APPLICANT: Jiang Yugu
: APPLICANT: Reed, Steven G.
: APPLICANT: Kalos, Michael
: APPLICANT: Fanger, Gary
: APPLICANT: Retter, Mark
: APPLICANT: Solk, John
: APPLICANT: Day, Craig

```

```

: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: FILE REFERENCE: 210121.427C9
: CURRENT APPLICATION NUMBER: US/09/439,313
: CURRENT FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 575
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 526
: LENGTH: 963
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-439-313-526

alignment_scores:
  Quality: 374.00      Length: 305
  Ratio: 1.861        Gaps: 9
  Percent Similarity: 65.902   Percent Identity: 30.492

alignment_block:
US-09-975-308-9 x US-09-439-313-526 ..
Align seg 1/1 to: US-09-439-313-526 from: 1 to: 963

9  pheIleIleLeuGlyLeuThrLysLysProGlnLeuGlnGlyIleLeuPh 25
   ::::::::::::::::::::
31  TTGTGCTTATTGGTATC.....CCAGATTAGAGAAAGCCCATTT 71
25  eLeuPhe.....PheLeuIleValTyrLeuValAlaPheLeuGlyA 39
   ::::::::::::::::::::
72  CTGGGTTGGCTTCCCTCCCTTTCATGATGATGGCATGTTTGGAA 121
39  smetLeuIleIleIleAlaLysIleTyrSerAsnThrLeuHisThrPro 55
   ::::::::::::::::::::
122  ACTGATGCTGCTCTTACGTAAGACGACGACGCTCCAGCTCCG 171
56  MetTyrValPheLeuLeuThrLeuAlaValAlaAspIleIleCysThrTh 72
   ::::::::::::::::::::
172  ATGATCCTTTCTCTGATGCTTGACACCATGACCTGAGCTTATCCAC 221
72  rserIleIleProLysMetLeuGlyThrMetLeuThrSerGlnAsnThrI 89
   ::::::::::::::::::::
222  ATCCACCAATGCTAGATCCCTTCTGTTGATTTGCCAGAGA 271
89  leSerTyrAlaGlyCysMetSerGlnLeuPheLeuPheThrTrpSerLeu 105
   ::::::::::::::::::::
272  TTAGCTTTGAGGCTGTCTTACCAGATGTC...TTATTCATGCCCTC 318
106  GlyAla...GlnMetValLeuPheThrThrMetAlaTyrAspArgTyrVa 121
   ::::::::::::::::::::
319  TGACCATTTGAATCCACCATCTGCTGGGCATGGCTTGAACGTTATGT 368
121  lAlaIleCysPheProLeuHisTyrSerThrIleMetLysHisHisMetC 138
   ::::::::::::::::::::
369  GGCCATCTCTCCACCATCTGCGCATGCTGCAGTGTCAACAT..... 411
138  ySValAlaLeuLeuSerMetValMetAlaIleAlaValThrAsnSerTrp 154
   ::::::::::::::::::::
412  ....ACAGTAACAGCCAGATTGGCATCTGCTGTGTCGGCGGATCC 456
155  ValHis.....ThrAlaLeuIleMetArgLeuThrPheCysGln 167
   ::::::::::::::::::::
457  CTTCTTTTTCCTCCACTGCTGCTGGCATGACAGCGGCTGCTGCGCA 506
167  yProAsnThrIleAspHisPhePheCysGlnIleProProLeuLeuAlaI 184
   ::::::::::::::::::::
507  CTCCAATGTCTCTCGCATCTCTATTTGTGCCACGAGTAAATGAACT 556
184  euserCysSerProValArgIleAsnGlnValMetValTyrValAlaAsp 200
   ::::::::::::::::::::
557  TGCCCTATGACAGACACTTGGCCAAATGTGTATATGTTTACTGCC... 603
201  IleThrLeuAlaIleGly...AspPheIleLeuThrCysIleSerTyrGl 216

```

LOCATION: 274..1233

Align seg 1/1 to: US-08-465-980-1 from: 1 to: 1474

```

29  leu1val1eul1eul1al1alpheu1eug1uysan1met1eul1e1le1le1 45
30  :::::::::::::::::::: |::|::| ::::::::::::::::::::
31  leu1val1eul1eul1al1alpheu1eug1uysan1met1eul1e1le1le1 45
32  :::::::::::::::::::: |::|::| ::::::::::::::::::::
33  364  ctttccatgatgatgatgtggcgaatgtgtggaaactgcattgtgtctcat 413
34  :::::::::::::::::::: |::|::| ::::::::::::::::::::
35  45  alu1le1tser1asn1thr1leu1n1ist1thr1promet1yval1phe1leu1t 62
36  :::::::::::::::::::: |::|::| ::::::::::::::::::::
37  414  cgtaaagaaacgaaacccactgcacagccgctgaaatgtaaccttcttgcga 463
38  :::::::::::::::::::: |::|::| ::::::::::::::::::::
39  62  hr1eul1al1val1al1as1p1le1le1cys1thr1thr1ser1le1le1pro1uys1met 78
40  :::::::::::::::::::: |::|::| ::::::::::::::::::::
41  464  tgccttcgaacccatttgaactcctggccttattgcacattccacattgccttaagatc 513
42  :::::::::::::::::::: |::|::| ::::::::::::::::::::
43  79  leu1g1t1hr1met1leu1thr1ser1gu1asn1thr1le1ser1tyr1al1g1cys1me 95
44  :::::::::::::::::::: |::|::| ::::::::::::::::::::
45  514  cttgccscattttctgtgttgattgattcccgagagaattagatattgagacgctgtct 563
46  :::::::::::::::::::: |::|::| ::::::::::::::::::::
47  95  tser1leu1eul1phe1leu1thr1trp1ser1leu1glu1ala1..glu1met1al1 111
48  :::::::::::::::::::: |::|::| ::::::::::::::::::::
49  564  taaccagaatgttc...tttatcagacccttcaccacattgaatcaccaca 610
50  :::::::::::::::::::: |::|::| ::::::::::::::::::::
51  111  eup1het1thr1met1ala1tyr1as1p1arg1tyr1val1al1e1cys1phe1pro1leu 127
52  :::::::::::::::::::: |::|::| ::::::::::::::::::::
53  611  ttcctctctggccatggcscctttagacscctttagtggccattcggcaccaccatcg 660
54  :::::::::::::::::::: |::|::| ::::::::::::::::::::
55  128  his1tyr1ser1thr1leu1asn1his1metcys1val1alal1eul1eul1ser1ne 144
56  :::::::::::::::::::: |::|::| ::::::::::::::::::::
57  661  cgccattgcctgcacgtcctcaacat.....acagtaaacagccca 698
58  :::::::::::::::::::: |::|::| ::::::::::::::::::::
59  144  tval1met1al1al1val1al1thr1asn1ser1tyr1al1his.....t 157
60  :::::::::::::::::::: |::|::| ::::::::::::::::::::
61  699  gat1tggcattcgtggctgtgtggctcggcgagactccttttttttcccaactgc 748
62  :::::::::::::::::::: |::|::| ::::::::::::::::::::
63  157  hr1alal1eul1eul1met1arg1leu1thr1phe1cys1glu1pro1asn1thr1leu1asp1is 173
64  :::::::::::::::::::: |::|::| ::::::::::::::::::::
65  749  cttcctcctgaacagcgctggccttctgcacattccacattgcctccttcggcac 798
66  :::::::::::::::::::: |::|::| ::::::::::::::::::::
67  174  phe1phe1cys1glu1le1pro1leu1eul1alal1eul1ser1cys1ser1pro1al1ar 190
68  :::::::::::::::::::: |::|::| ::::::::::::::::::::
69  799  tcc1aat1ttgtgtccacacagatgataatgaagtggcctatgcagacacattt 848
70  :::::::::::::::::::: |::|::| ::::::::::::::::::::
71  190  g1le1asn1glu1met1val1tyr1val1al1as1p1le1le1hr1leu1al1al1e1glu1 206
72  :::::::::::::::::::: |::|::| ::::::::::::::::::::
73  849  gcccaatgtgtaattgatgtcttactgcgc...atttcgtcgtgcacagggcg 895
74  :::::::::::::::::::: |::|::| ::::::::::::::::::::
75  207  .asp1phe1leu1thr1cys1le1ser1tyr1glu1phe1le1le1val1al1a1le 222
76  :::::::::::::::::::: |::|::| ::::::::::::::::::::
77  896  tggcagatgatacttcattcctcttgccttatttctgataatagacagcgtt 945
78  :::::::::::::::::::: |::|::| ::::::::::::::::::::
79  223  leu1arg1le1arg1thr1val1glu1c1y1s1arg1ys1al1a1phe1ser1thr1cys1se 239
80  :::::::::::::::::::: |::|::| ::::::::::::::::::::
81  946  ctgcacatgccttccaaactcagaacggcgccaaaggcctttggaaactgtgt 995
82  :::::::::::::::::::: |::|::| ::::::::::::::::::::
83  239  tser1his1leu1thr1val1al1thr1leu1tyr1tyr1ser1pro1al1le1tyr1thr1t 256
84  :::::::::::::::::::: |::|::| ::::::::::::::::::::
85  996  gtccacacattgatgtgtggtgacacgaccttctatgtgcacattattggcgcctt 10454
86  :::::::::::::::::::: |::|::| ::::::::::::::::::::
87  256  ty1le1.....arg1pro1al1ser1ser1tyr1thr1phe1glu1arg1asp1y1s1al1 270
88  :::::::::::::::::::: |::|::| ::::::::::::::::::::
89  1046  cag1ttgtacacccgcttttggaaaacaccccttcatccattgcctgctgattgttc 10955

```



```

APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAL70
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARILLA, BYRNE, BAIN, GILFILLAN, CECCHI,
ADDRESSEE: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07093
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,980
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
PCT-US95-07093-1

alignment_scores:
Quality: 363.00 Length: 282
Ratio: 1.931 Gaps: 7
Percent Similarity: 66.667 Percent Identity: 31.206

alignment_block:
US-09-975-308-9 x PCT-US95-07093-1 ..

Align seg 1/1 to: PCT-US95-07093-1 from: 1 to: 1474

29 LeuILEValTYrLeuValAlaPheLeuGLYAsnMetLeuILElleIleAl 45
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 CTTTCATGATGATGATGCAATGTGGAACATGCAATGCTGCTCAT 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 ALYSILETYrSerAsnThrLeuHisThrPrometTYrValPheLeuLeuT 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 CGTAGAGACGACGACGACGCTGACGCTCGCATGATCTCTTCTGCA 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 hrLeuAlaValAlaAspIleIleCysThrThrSerIleIleProLYsMet 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 TGCCTTGAGCATGACCTGCTTATCCACATCCACCATGCTTATCATC 513
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 LeuGLYThrMetLeuThrSerGLYAsnThrIleSerTYrAlaGLYcysMe 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 CTTGCGCTTTTCGTTGATTCGAGAGATTAATGACATGAGCGCTCTCT 563
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 tSerGLYLeuPheLeuPheThrTrpSerIleuGLYAla...GluMetValL 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 TACCACAGATGTTTC...TTTATTCATGCGCTCTGACCATTAATGCCACA 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

111 eupheThrThrMetAlaTYrAspArgTYrValAlaIleCysPheProLeu 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 TCCTGTGGCCATGCGCTTGTACGTTATGTGGCATGTGCGACCATG 660
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 HisTYrSerThrIleMetAsnHisHisMetCysValAlaLeuLeuSerMe 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 CGGCATGCTGACGTGCTCAACAT.....ACAGTAACAGGCCA 698
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
144 tValMetAlaIleAlaValThrAsnSerTYrValHis.....T 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 GATTGGCATGCTGCTGTGTCGCGGATCCCTCTTTTTCCTCCATGCG 748
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 hrAlaLeuIleMetArgLeuThrPheCysGLYProAsnThrIleAspHis 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
749 CTCGTGTGATTAACGGCTGCGCTTCTGACACTCAATGCTCTCTGAC 798
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 PhePheCysGLYIleProProLeuLeuAlaLeuSerCysSerProValr 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
799 TCCTATGTGTCCACACAGATGTATGAGTTGGCTATGACAGACATTT 848
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 gILAsnGLYValMetValTYrValAlaAspIleThrLeuAlaIleGLY. 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
849 GCCCAATGTGTATATGCTTACTGCC...ATTCTGCTGCTCATGGCG 895
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 .AspPheIleLeuThrCysIleSerTYrGLYPheIleIleValAlaIle 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
896 TGGACGTATATTCATCTCTGCTGCTATTTTCGATATATCGAACGTT 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 LeuArgIleArgThrValGLYGLYsArgLYsIaPheSerThrCysSe 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 CTGCACATGCTCTCCAGTCAAGCGGACGCGCAAGCTTTGGAACTGTGT 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 rSerHisLeuThrValAlaThrLeuTYrTYrSerProValIleTYrThr 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 GTCACACATGTGTGTGCTGCTGCTCTGCTATGTGCTATTTGGCTCT 1045
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 yTlle.....ArgProAlaSerSerTYrThrPheGLYArgAspLYsVal 270
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1046 CAGTTGTACACCGCTTGGAAACGCTTCATCCCATTTGCGGTGTC 1095
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 ValAlaAlaLeuTYrThrLeuValTYrProThrLeuAsnPrometValTY 287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1096 ATGGGTGACATCTACCTGCTGCTGCTGCTGCTGCTATCATCCATCACTA 1145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 rSerPheGLYAsnArgGLYMetGLYAlaGLYIleArgLYsValPhe 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1146 TGTGTCCAAAACCAACAGATCAACACAGGGGTGCTGCTATGTGTC 1191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: /cgn2_6/plodata/1/lna/5A_COMB.seq:US-08-671-525B-5

seq_documentation_block:
; Sequence 5, Application US/08671525B
; Patent No. 5703220
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671,525B
; FILING DATE: June 27, 1996

```

655 TACTCGAGACGAAAAATGGTCATTGTGTGCTCCATTCACCATCATGTTCTTCGAC 704
145 IMetAlaIleAlaValAlaThrAsnSerTrpValAlaHisThrAlaLeuIleMetA 162
705 CAtGATGCTCTCCATGATGGACCCCTCTACGAGCACAATGTTCTCTTGGCCG 754
162 rglEurThrPheCysGlyProAsnThrIleAspIshPhe.....PheCys 176
755 GGCTG.....CACGTCAAGGCCATACGA 777
177 GluIleProIleuAlaLeuSerCysSerProValArgIleAsnGI 193
778 GCATCGCACCT.....GCCAGGGGGTGGCCCAACAGCAA...CACATC 818
193 vAlaMetValTrpAlaAlaAspIleThrIleuAlaIleGlyAspPheIle 210
819 ATGCATGAAGGGGGCAGCACAACCATCACTTCCTCGGGGTGTTCACTT 868
210 eurThrCysIleSerTrpGlyPheIleIleValAlaIleLeuArgIleArg 226
869 TCTGCTGGCCCCCTTCCTCTCCACTGCTGCCTCATCATC..... 909
227 ThrValGluGlyLysArgLysAlaPheSerThrCysSerSerHisLeuTh 243
910ACCTGCCCCACAC..... 921
243 rValValThrLeuTrpTrpSerProValIleTrpThrIleArgProA 260
922ACCCCTACTGCATTCGCTACACACTGCCCACT 952
260 lAsrSerTrpThrPheGluArgAspLysValAlaAlaLeuTrpThr 276
953 TCACACACTAC.....CTGTCCTCATC 975
277 LeuValAlaProThrLeuAsnProMetValTrpSerPheGluAsnArgGI 293
976 ATGTGCAACTCCGTCAATTCACCCACTCACTACAGCTTCCGAGACCTGSA 1022
293 uMetGlnAlaGlyIleArgLysVal 301
1026 ATTGGCGAACACCTTTAAGGAGAGTT 1050

OM of: US-09-975-308-9 to: GenEmbl: * out_format : pfs
Date: Jun 24, 2002 12:54 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+g2n model -DEV=x1h  
-O=/cgn2_1/USPFO.spool/US0975308/runat_24062002_090303_26237/app-query.fasta-1.368  
-DB=GenEmbl -GPM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500  
-OGAPEXT=0.050 -XGAPEXT=10.000 -XGAPEXT=0.500 -FGAPEXT=6.000  
-FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DEL0P=6.000  
-DEL0P=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US0975308 @CGN1.1.9449  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPPX -WAIT -THREADS=1
```

Search information block:

Query: US-09-975-308-9
Query length: 307
Database: GenEmbl: *
Database sequences: 1797656
Search length: 187333701
Search time (sec): 1843.850000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_hhg:AC091612	- 1575.00	2435.56	2.7e-127	180657	AC091612 Homo sapiens chromo
gb_pat:AX241862	- 1574.00	2430.56	8.3e-130	921	AX241862 Sequence 610 from Pat
gb_hhg:AC026038	+ 1574.00	2436.33	2.4e-127	138591	AC026038 Homo sapiens chromo
gb_hhg:AL57039	+ 1574.00	2433.50	3.5e-127	190889	AC057039 Homo sapiens chromo
gb_hhg:AC094718	- 1395.00	2152.23	1.6e-111	160483	AC094718 Rattus norvegicus c
gb_hhg:AC096601	- 1380.00	2125.69	4.9e-110	221466	AC096601 Mus musculus clone
gb_ro:AV073800	+ 1354.00	2132.95	1.9e-110	926	AV073800 Mus musculus olfacto
gb_pat:AX350633	+ 1339.00	2110.56	3.4e-109	798	AX350633 Sequence 53 from Pat
gb_pat:AX350675	+ 1443.00	1802.54	4.9e-92	663	AX350675 Sequence 11 from Pat
gb_hhg:AC101272	+ 954.50	1480.04	4.5e-74	59688	AC101272 Mus musculus clone
gb_ro:AV073135	+ 965.00	1503.47	2.2e-75	915	AV073135 Mus musculus olfacto
gb_ro:AV074180	+ 938.50	1476.61	6.9e-74	916	AV074180 Mus musculus olfacto
gb_hhg:AC094870	- 935.00	1428.61	6.2e-71	112730	AC094870 Rattus norvegicus c
gb_hhg:AC094700	- 935.00	1423.67	6.2e-71	137438	AC094700 Rattus norvegicus c
gb_hhg:AC096461	+ 906.00	1377.56	2.3e-68	204143	AC096461 Rattus norvegicus c
gb_ro:AV073132	+ 900.00	1415.65	1.7e-70	930	AV073132 Mus musculus olfacto
gb_ro:AV073176	+ 893.00	1404.59	7.1e-70	930	AV073176 Mus musculus olfacto
gb_ro:AV073062	+ 883.00	1388.79	5.4e-69	930	AV073062 Mus musculus olfacto
gb_hhg:AC094493	+ 873.00	1330.36	9.7e-66	116603	AC094493 Rattus norvegicus c
gb_ro:AV073127	+ 871.00	1369.35	6.4e-68	960	AV073127 Mus musculus olfacto
gb_ro:AV073133	+ 869.00	1366.61	9.3e-68	936	AV073133 Mus musculus olfacto
gb_pat:AX201672	+ 861.00	1351.49	6.5e-67	1241	AX201672 Sequence 30 from Pat
gb_hhg:AC011879	- 861.00	1308.58	1.6e-64	160654	AC011879 Homo sapiens clone
gb_pat:AX512324	- 861.00	1308.16	1.7e-64	168473	AX512324 Human DNA sequence
gb_pat:AX241531	+ 860.00	1352.51	3.7e-67	924	AX241531 Sequence 279 from Pat
gb_hhg:AC094700	+ 858.50	1302.81	3.3e-64	197438	AC094700 Rattus norvegicus c
gb_ro:AV073126	+ 857.00	1347.63	1.1e-66	939	AV073126 Mus musculus olfacto
gb_ro:AV073125	+ 852.00	1339.76	2.9e-66	936	AV073125 Mus musculus olfacto
gb_ro:AV073128	+ 851.00	1338.15	3.6e-66	939	AV073128 Mus musculus olfacto
gb_ro:AV074052	+ 850.50	1337.46	3.9e-66	928	AV074052 Mus musculus olfacto
gb_ro:AV073124	+ 850.00	1335.60	4.4e-66	936	AV073124 Mus musculus olfacto
gb_hhg:AC094870	+ 840.50	1279.31	6.8e-63	112730	AC094870 Rattus norvegicus c
gb_ro:AV073154	+ 839.00	1319.02	4.2e-65	957	AV073154 Mus musculus olfacto
gb_ro:AV073578	+ 835.50	1313.80	8.1e-65	924	AV073578 Mus musculus olfacto
gb_hhg:AC094493	- 835.50	1313.66	8.3e-65	939	AV073134 Mus musculus olfacto
gb_ro:AV073577	- 830.50	1263.22	5.3e-62	116603	AC094493 Rattus norvegicus c
gb_ro:AV073073	+ 816.50	1283.78	3.8e-63	924	AV073577 Mus musculus olfacto
gb_hhg:AC108568	+ 755.00	1186.54	9.9e-58	933	AV073073 Mus musculus olfacto
gb_hhg:AC108568	+ 750.00	1137.77	5.2e-55	95794	AC108568 Rattus norvegicus c
gb_ro:AV073818	+ 744.00	1168.88	9.6e-57	963	AV073818 Mus musculus olfacto

gb_ro:AY073328 + 740.00 1162.78 2.1e-56 939 AY073328 Mus musculus olfact
gb_ro:AY073679 + 732.50 1150.85 9.7e-56 948 AY073679 Mus musculus olfact
gb_ro:AY073680 + 731.50 1149.21 1.2e-55 954 AY073680 Mus musculus olfact
gb_hhg:AC095959 - 731.50 1101.03 5.8e-53 224606 AC095959 Mus musculus c10
gb_ro:AY073249 + 730.00 1146.98 1.6e-55 939 AY073249 Mus musculus olfact

seq_name: gb_hhg:AC091612

seq_documentation_block:

LOCUS AC091612 180657 bp DNA linear HTG 05-FEB-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-656022, WORKING DRAFT
SEQUENCE 1 unordered pieces.
AC091612 AL390860
VERSION AC091612.4 GT:18497169
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 180657)
Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
Unpublished
2 (bases 1 to 180657)
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and
Haugen, E.D.
Direct Submission
Submitted (09-MAY-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Feb 5, 2002 this sequence version replaced gl:15487406.

TITLE

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
Project Information
Center project name: chr-1
Center clone name: RP11-656022 (sc0182)
----- Summary Statistics
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180536 bases at least Q40
Consensus quality: 180557 bases at least Q20
Insert size: 194815; 11.0% error; agarose-fp
Insert size: 180657; sum-of-contigs
Quality coverage: 8.4x in Q20 bases; agarose-fp
Quality coverage: 9.0x in Q20 bases; sum-of-contigs

COMMENT

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 180657: contig of 180657 bp in length.
Location/Qualifiers
1. 180657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-656022"
/clone_lib="RPCI human BAC library 11"
1. 180657
/note="assembly name:Contig19"
BASE COUNT 52363 a 34237 c 35512 g 58545 t
ORIGIN

FEATURES

source
misc_feature
/note="assembly name:Contig19"

BASE COUNT

52363 a 34237 c 35512 g 58545 t

ORIGIN

alignment_scores:

Quality: 1575.00 Length: 307
 Ratio: 5.130 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.674

alignment_block:

US-09-975-308-9 x AC091612/rev ..

Align seg 1/1 to reverse of: AC091612 from: 1 to: 180657

```

1 MetAsnHisSerValValThrGluPheIleIleuGlyLeuThrLys 17
  |||
155813 ATGATACACAGCGTTGTAAGTCAATTAATTTGGGCTCCACCAAAA 155764

17 sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
  |||
155763 GCCGCACTCCAGGGAATTAATCTTCCTTTTCTCATGTCTATCTG 155714

34 aAlaIlePheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsn 50
  |||
155713 TGGCTTTTCTCGGCAACATGCTCATCATTCGCAAAATCTATTAACAA 155664

51 ThrLeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAs 67
  |||
155663 ACCTTGCGATACGCCCATATATGTTTCTTCTGACACTGGCTGTGGGA 155614

67 pIleIleCysThrThrSerIleIleIleProLysMetLeuGlyThrMetLeu 84
  |||
155613 CATCATCTGCACAAACACATCATACCGAAGATGCTGGGACCATGCTAA 155564

84 hrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGluPheLeu 100
  |||
155563 CATCGAAATAACATTTCATATGACAGGCTGCATGCCAGCTCTCTTG 155514

101 PheThrTrpSerLeuGlyAlaGluMetValLeuPheThrMetAlaTy 117
  |||
155513 TTCACATGAGTCTCTGGAGCTGAGATGGTCTCTCACACCATGGCTTA 155464

117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIleMet 134
  |||
155463 TGACCGCATATGGCCATTGTTCCCTTCATTAACAGTACTATTATGA 155414

134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
  |||
155413 ACCACCATATGATGTAGAGCTTGCTGACAGATGCTATGCTATGACAG 155364

151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysG 167
  |||
155363 ACCAATTCCTGGGTGACACAGCTCTTATCATGAGGTGACTTCTGTGG 155314

167 yProAsnThrIleAspHisPhePheCysGluLeuProPheLeuAlaL 184
  |||
155313 GCCAAACACCATTTGACACTTCTTCTGTGAGATACCCCATTTGCTGG 155264

184 euSerCysSerProValArgIleAsnGluValMetValTyrValAlaAsp 200
  |||
155263 TGTCCGTGAGCCCTGTAAAGATCAATGAGGTGATGTATGTCTGTGAT 155214

201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGly 217
  |||
155213 ATTAACCTCGCCCAATGAGGCACTTAATCTTACCTGCATCTCCATGCTTT 155164

217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLys 234
  |||
155163 TATCATGTTGTCATATCTCCGTATCCGCAACAGTAAGAGGAGAGAGAG 155114

234 laPheSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSer 250
  |||
155113 CCTTCACACATGCTCATCTCATCTCACAGTGTGACCCCTTACTATTC 155064

251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlu 267
  |||
155063 CCGTATATCTACACGTATATCCGCCCTGCTTCCAGCTATACATTGAAG 155014

```

267 gaPheLysValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsn 284

155013 AGCAAGAGGTGAGTGCATCTCTATCTCTGTGATCTCCACATTTAAAC 154964

284 rOmEValTyrSerPheGlnAsnArgLumetGlnAlaGlyIleArgLys 300

154963 CGATGCTGTACACCTTCCAGAAATAGGAGATGCAGCAGGAAATTAGGAAG 154914

301 ValPheAlaPheLeuLysHis 307

154913 GTGTTGCAATTTCTGAACAC 154893

seq_name: gb_pat:AX241862

seq_documentation_block: 921 bp DNA linear PAT 26-SEP-2001

LOCUS AX241862 Sequence 610 from Patent WO0127158.

DEFINITION AX241862

ACCESSION AX241862

VERSION AX241862.1 GI:15798737

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

alignment_scores:

Quality: 1574.00 Length: 307
 Ratio: 5.127 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.349

alignment_block:

US-09-975-308-9 x AX241862 ..

Align seg 1/1 to: AX241862 from: 1 to: 921

```

1 MetAsnHisSerValValThrGluPheIleIleuGlyLeuThrLys 17
  |||
1 ATGAATCAACAGCGTTGTAAGTCAATTAATTTGGGCTCCACCAAAA 50

17 sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
  |||
51 GCCTAACATCCAGGGAATTAATCTCTCTTTTCTCATGTCATATCTG 100

34 aAlaIlePheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsn 50
  |||
101 TGGCTTTTCTCGGCAACATGCTCATCATTCATTTGCCAAATCTATTAACAA 150

51 ThrLeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAs 67
  |||
151 ACCTTGCGATACGCCCATATATGTTTCTTCTGACACTGGCTGTGGGA 200

67 pIleIleCysThrThrSerIleIleIleProLysMetLeuGlyThrMetLeu 84
  |||
201 CATCATCTGCACAAACAGCATCATACCGAAGATGCTGGGACCATGCTAA 250

84 hrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGluPheLeu 100
  |||
251 CATCAAGAAATACCATTTATATGACAGGCTGCATGTGCCAGCTCTTCTTG 300

```

101 PheHrTrpSerLeuGlyAlaGluMetValLeuPheHrThrMetAlaTy 117
 301 TTGCATGAGTCTCTGGAGCTGAGATGTTCTCTTCCACACATGCGCTTA 350
 117 rAspArgTyValAlaIleCysPheProLeuHisTySerThrIleMetAla 134
 351 TGACCGCTATGTGGCCATTGTTCCCTCTTCATATACAGTACGTATGA 400
 134 smHSHSmCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
 401 ACCACCATATGTTGTAGCTCTGCTCAGCATGTCATGCTATTTGCAATC 450
 151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysG 167
 451 ACCAATTCCTGGGTGCACACAGCTTTATATATGATGAGTTGCTTCTGG 500
 167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAla 184
 501 GCCAAGACCATGACCATTTCTGTGAGATACCCCATGCTGCTGCTT 550
 184 euserCysSerProValArgIleAsnGluValMetValTyrrValAlaAsp 200
 551 TGTCGTGTACCCCTGTAGATCATGATGATGATGATGATGATGATGAT 600
 201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrglyph 217
 601 ATTACCTGGCCATAGGGGACTTATTTCTTACTGCTATCCTATGTTT 650
 217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysA 234
 651 TATCATGTTGCTATCTCCGTATCCGACAGTACAGTACAGTACAGTACAG 700
 234 lAheSerTrpCysSerSerHisLeuThrValValThrLeuTyrrTyrrSer 250
 701 CCTTCTCAATGCTCATCTCATCTCATCTCATCTCATCTCATCTCATCT 750
 251 ProValIleTyrrThrTyrrIleArgProAlaSerSerTyrrThrPheGlu 267
 751 CCTGTATCTACACCTATATCCGCTGCTTCCAGCATATCATTTGAAG 800
 267 gAspLysValValAlaIleLeuTyrrThrLeuValThrProThrLeuAsn 284
 801 AGCAAGGAGGTGAGTGCACACTCTTGTGACTCCACATTTAAACC 850
 284 rOmetValTyrrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300
 851 CGATGCTGTACACCTTCCAGATATGAGGATGAGGATGAGGATGAGGAT 900
 301 ValPheAlaPheLeuLysHis 307
 901 GTGTTGCAATTTCTGAACAC 921
 seq_name: gb_hhg:AC026038
 seq_documentation_block:
 LOCUS AC026038 138591 bp DNA linear HTG 01-SEP-2000
 DEFINITION Homo sapiens chromosome 1 clone RP11-109C14, WORKING DRAFT
 SEQUENCE 22 unordered pieces.
 ACCESSION AC026038.4 GI:9958133
 VERSION AC026038
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 138591)
 TITLE Waterston, R.H.
 JOURNAL The sequence of Homo sapiens clone
 REFERENCE 2 (bases 1 to 138591)
 JOURNAL Unpublished
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2000) Genome Sequencing Center, Washington

COMMENT

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 1, 2000 this sequence version replaced gi:8567954.
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0109C14
 Summary Statistics -----
 Sequencing vector: p3: 100%
 Chemistry: Dye-terminator; 100% of reads
 Chemistry: Dye-terminator; Big Dye; 0% of reads
 Assembly program: Phrap; Version 0.990319
 Consensus quality: 129539 bases at least Q40
 Consensus quality: 131919 bases at least Q30
 Consensus quality: 13309 bases at least Q20
 Insert size: 14000; agarose-fp
 Insert size: 136491; sum-of-contigs
 Quality coverage: 3.88 in Q20 bases; agarose-fp
 Quality coverage: 4.04 in Q20 bases; sum-of-contigs
 ----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 5221: contig of 5221 bp in length
 * 5222 5321: gap of unknown length
 * 5322 9571: contig of 4250 bp in length
 * 9572 9671: gap of unknown length
 * 9672 13673: contig of 4002 bp in length
 * 13674 13773: gap of unknown length
 * 13774 18721: contig of 4948 bp in length
 * 18722 18821: gap of unknown length
 * 18821 24023: contig of 5202 bp in length
 * 24024 24123: gap of unknown length
 * 24124 28988: contig of 4865 bp in length
 * 28989 29088: gap of unknown length
 * 29089 34171: contig of 5033 bp in length
 * 34172 34271: gap of unknown length
 * 34272 40989: contig of 6718 bp in length
 * 40990 41089: gap of unknown length
 * 41090 48355: contig of 7266 bp in length
 * 48356 48455: gap of unknown length
 * 48456 59366: contig of 10911 bp in length
 * 59367 59466: gap of unknown length
 * 59467 61313: contig of 1847 bp in length
 * 61314 61413: gap of unknown length
 * 61414 71341: contig of 9828 bp in length
 * 71342 71342: gap of unknown length
 * 71343 80867: contig of 9526 bp in length
 * 80868 80967: gap of unknown length
 * 80968 96688: contig of 18721 bp in length
 * 96689 99788: gap of unknown length
 * 99789 122864: contig of 23076 bp in length
 * 122865 122964: gap of unknown length
 * 122965 124804: contig of 1840 bp in length
 * 124805 124904: gap of unknown length
 * 124905 126520: contig of 1716 bp in length
 * 126521 126720: gap of unknown length
 * 126721 128799: contig of 2079 bp in length
 * 128800 128899: gap of unknown length
 * 128900 130561: contig of 1662 bp in length
 * 130562 130651: gap of unknown length
 * 130652 132547: contig of 1886 bp in length
 * 132548 132647: gap of unknown length
 * 132648 135428: contig of 2781 bp in length

* 135428 135528: gap of unknown length
 * 135529 138591: contig of 3063 bp in length.
 FEATURES
 Location/Qualifiers
 1. 138591
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-109C14"
 1. 5221
 /note="assembly_name:Contig10
 clone_end:77
 vector_side:left"

misc_feature
 /note="assembly_name:Contig11"
 5322..9571
 misc_feature
 /note="assembly_name:Contig12"
 9672..13673
 misc_feature
 /note="assembly_name:Contig13"
 13774..18721
 misc_feature
 /note="assembly_name:Contig13"
 18822..24023
 misc_feature
 /note="assembly_name:Contig14"
 24124..28988
 misc_feature
 /note="assembly_name:Contig15"
 29089..34171
 misc_feature
 /note="assembly_name:Contig16"
 34272..40989
 misc_feature
 /note="assembly_name:Contig17"
 41090..48355
 misc_feature
 /note="assembly_name:Contig18"
 48456..59366
 misc_feature
 /note="assembly_name:Contig19"
 59467..61313
 misc_feature
 /note="assembly_name:Contig2"
 61414..71241
 misc_feature
 /note="assembly_name:Contig20"
 71342..80867
 misc_feature
 /note="assembly_name:Contig21"
 80968..93688
 misc_feature
 /note="assembly_name:Contig22"
 99789..122864
 misc_feature
 /note="assembly_name:Contig23
 clone_end:SP6
 vector_side:left"
 122965..124804
 misc_feature
 /note="assembly_name:Contig3"
 124905..126520
 misc_feature
 /note="assembly_name:Contig4"
 126721..128799
 misc_feature
 /note="assembly_name:Contig5"
 128900..130561
 misc_feature
 /note="assembly_name:Contig6"
 130662..132547
 misc_feature
 /note="assembly_name:Contig7"
 132648..135428
 misc_feature
 /note="assembly_name:Contig8"
 135529..138591
 misc_feature
 /note="assembly_name:Contig9"
 25199 a 25318 c 25318 g 41754 t 2109 others
 BASE COUNT 44211

alignment_scores:
 Quality: 1574.00 Length: 307
 Ratio: 5.127 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.349

alignment_block:
 US-09-975-308-9 x AC026038 ..

Align seg 1/1 to: AC026038 from: 1 to: 138591

1 MetcshntsservAlaValThrGluPheIleIleuGlyLeuThrIleuVal
 |||||||
 86786 ATGAATACACAGCGTTGTAAGTCAATTAATTTCTGGGCTCACCAAAAA 86835

17 sProGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrlleuVal 34
 |||||||
 86836 GCGCTACATCCAGGAAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 86885
 34 aLaPheLeuGlyAsnMetLeuIleIleIleValValIleTyrlSerAsn 50
 |||||||
 86886 TGGCTTTCTGGGCAACATGCTCATCATCATGCGCAAAATCATTAACAC 86935
 51 ThrLeuHisThrProMetTyrlValPheLeuThrLeuAlaValAlaVal 67
 |||||||
 86936 ACCTTGACATACGCCCATGATGTTTCTCTGACACCTGCTGTTTGA 86985
 67 PheIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeu 84
 |||||||
 86986 CATCATCTGCACACACAGCATCATACCGAAGATGCTGGGACCATGCTAA 87035
 84 hSerGluAsnThrIleSerTyrlAlaGlyCysMetSerGluLeuPheLeu 100
 |||||||
 87036 CATCAGAAATACCATTTCTATGACAGGCTGCATGCTCCAGCTCTTCTG 87085
 101 PheThrTrpSerLeuGlyValAlaGluMetValLeuPheThrThrMetAlaTy 117
 |||||||
 87086 TTCACATGCTCTCTGGACCTGAGATGATGTTCTTCCACACCATGAGCTTA 87135
 117 rAspArgTyrlValAlaIleCysPheProLeuHisTyrlSerThrIleMetAl 134
 |||||||
 87136 TGACCGCTATGCGCATTTGTTCCCTTCATTAACATGATGTTATGA 87185
 134 snHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
 |||||||
 87186 ACCACCATATGATGTAGCTTCTGTCACAGATGCTCATGCTATGACAGTC 87235
 151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlu 167
 |||||||
 87236 ACCAATTCCTGGGTCACACAGCTTATCATGAGTTGATCTTCTGTGG 87285
 167 yProAsnThrIleAspHisPheCysGluIleProLeuLeuAlaIle 184
 |||||||
 87286 GCCAAACACCATGACCATCTTCTGTGATACCCCATCTCTGCTT 87335
 184 euserCysSerProValArgIleAsnGluValMetValTyrlValAlaAsp 200
 |||||||
 87336 TGTCCTGTAGCCCTGTAAGATCAATGAGTGAATGATGTTGTTGCTGAT 87385
 201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrlGlyP 217
 |||||||
 87386 ATTAACCTGGCCATAGGAGCTTATCTTACCTGCAATCTCTATGAGTTT 87435
 217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysA 234
 |||||||
 87436 TATCATTTGTTCTATTCCTCCATCCGACAGTAGAAGGCAAGAGAAAG 87485
 234 lApeSerThrCysSerSerHisLeuThrValAlaIleThrLeuTyrlSer 250
 |||||||
 87486 CCTTCTCAACATGCTCATCTCATCTCACAGTGAGGACCTTATCATTTCT 87535
 251 ProValIleTyrlThrTyrlLeuArgProAlaSerSerTyrlThrPheGluAr 267
 |||||||
 87536 CCTGTAATCTACCTATATCCGCCCTCTCTCCAGCATATATGAAAG 87585
 267 gAspLysValValAlaAlaLeuTyrlThrLeuValThrProThrLeuAsn 284
 |||||||
 87586 AGACAAAGGTGAGTGCACCTATATCTTGTGACTCCACATTAAC 87635
 284 rOmetValTyrlSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLys 300
 |||||||
 87636 CGATGCTGTAAGCTTCCAGATAGGAGATGACGACGAGATTTAGGAAG 87685
 301 ValPheAlaPheLeuLysHis 307
 |||||||
 87686 GTGTTTGCATTTCTGAACAC 87706

seq_name: gb_htg:AL57039

seq_documentation_block: 190889 bp DNA linear HTG 19-SEP-2001

LOCUS AL57039

DEFINITION Homo sapiens chromosome 1 clone RP11-634B7, *** SEQUENCING IN PROGRESS ***, 18 unordered pieces.

ACCESSION AL57039

VERSION AL57039.5 GI:9797648

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 190889)

Plumb,B.

Direct Submission

Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

Requests: clonerequests@sanger.ac.uk

On Aug 12, 2000 this sequence version replaced gi:9214044.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: ba634b7

----- Summary Statistics

Assembly program: XGAP; version 4.5

Sequencing vector: plasmid; 108752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 183408 bases at least Q40

Consensus quality: 186557 bases at least Q30

Consensus quality: 188087 bases at least Q20

Insert size: 189189; sum-of-contigs

Insert size: 188721; 6.6% error; agarose-fp

Quality coverage: 4.17x in Q20 bases; sum-of-contigs Quality coverage: 4.30x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 12042: contig of 12042 bp in length

* 12043 12142: gap of 100 bp

* 12143 30262: contig of 18120 bp in length

* 30263 30362: gap of 100 bp

* 30363 41423: contig of 11063 bp in length

* 41426 41525: gap of 100 bp

* 41526 48351: contig of 6826 bp in length

* 48352 48451: gap of 100 bp

* 48452 57808: contig of 9357 bp in length

* 57809 57908: gap of 100 bp

* 57909 68643: contig of 10735 bp in length

* 68644 68743: gap of 100 bp

* 68744 75536: contig of 6793 bp in length

* 75537 75636: gap of 100 bp

* 75637 84190: contig of 8554 bp in length

* 84191 84290: gap of 100 bp

* 84291 88039: contig of 4749 bp in length

* 89040 89139: gap of 100 bp

* 89140 94516: contig of 5377 bp in length

* 94517 94616: gap of 100 bp

* 94617 105920: contig of 11304 bp in length

* 105921 106020: gap of 100 bp

* 106021 133564: contig of 27544 bp in length

* 133565 137940: gap of 100 bp

* 137941 138040: contig of 4276 bp in length

* 137941 138040: gap of 100 bp

* 138041 141386: contig of 3346 bp in length

* 141387 141486: gap of 100 bp

* 141487 143513: contig of 2027 bp in length

* 143514 143613: gap of 100 bp

* 143614 159175: contig of 15562 bp in length

* 159176 159275: gap of 100 bp

* 159276 174409: contig of 15134 bp in length

* 174410 174509: gap of 100 bp

* 174510 190889: contig of 16380 bp in length.

----- Location/Qualifiers

1..190889

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-634B7"

/clone_1fb="RPC1-11.3"

1..12042

/note="assembly_fragment:01421"

fragment_chain:1

clone_end:77

vector_side:left

12143..30262

/note="assembly_fragment:00925"

fragment_chain:1

30363..41425

/note="assembly_fragment:00451"

fragment_chain:2

41326..48351

/note="assembly_fragment:00297"

fragment_chain:2

48452..57808

/note="assembly_fragment:00082"

fragment_chain:2

57909..68643

/note="assembly_fragment:01434"

fragment_chain:2

68744..75536

/note="assembly_fragment:01941"

fragment_chain:2

75637..84190

/note="assembly_fragment:00841"

fragment_chain:2

84291..89039

/note="assembly_fragment:00517"

fragment_chain:3

89140..94516

/note="assembly_fragment:00194"

fragment_chain:3

94617..105920

/note="assembly_fragment:00685"

fragment_chain:4

106021..113564

/note="assembly_fragment:00244"

fragment_chain:4

113665..137940

/note="assembly_fragment:00993"

fragment_chain:5

138041..141386

/note="assembly_fragment:00049"

fragment_chain:5

141487..143513

/note="assembly_fragment:01137"

143614..159175

/note="assembly_fragment:01189"

159276..174409

/note="assembly_fragment:01199"

174510..190889

/note="assembly_fragment:01519"

BASE COUNT 60637 a 34972 c 35643 g 57919 t 1712 others

ORIGIN

alignment_scores:

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624554.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRFP

Center clone name: CH230-5M10

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to findPhrapList

Consensus quality: 127493 bases at least Q40

Consensus quality: 137458 bases at least Q30

Consensus quality: 144882 bases at least Q20

Estimated insert size: 129461; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 65 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 5754: contig of 5754 bp in length

* 5755 5854: gap of unknown length

* 5855 11490: contig of 5636 bp in length

* 11491 11590: gap of unknown length

* 11591 17216: contig of 5626 bp in length

* 17217 17316: gap of unknown length

* 17317 21754: contig of 4438 bp in length

* 21755 21854: gap of unknown length

* 21855 25351: contig of 3497 bp in length

* 25352 25451: gap of unknown length

* 25452 30651: contig of 5200 bp in length

* 30652 30751: gap of unknown length

* 30752 34861: contig of 4110 bp in length

* 34862 34961: gap of unknown length

* 34962 38499: contig of 3538 bp in length

* 38500 38599: gap of unknown length

* 38600 41784: contig of 3185 bp in length

* 41785 41884: gap of unknown length

* 41885 45873: contig of 3989 bp in length

* 45874 45973: gap of unknown length

* 45974 50324: contig of 4251 bp in length

* 50325 50324: gap of unknown length

* 50325 53266: contig of 2942 bp in length

* 53267 53366: gap of unknown length

* 53367 56430: contig of 3064 bp in length

* 56431 56530: gap of unknown length

* 56531 59835: contig of 3305 bp in length

* 59836 59936: gap of unknown length

* 59936 62863: contig of 2928 bp in length

* 62864 62963: gap of unknown length

* 62964 65355: contig of 2392 bp in length

* 65356 65455: gap of unknown length

* 65456 68305: contig of 2850 bp in length

* 68306 68405: gap of unknown length

* 68406 70717: contig of 2312 bp in length

* 70718 70817: gap of unknown length

* 70818 73306: contig of 2489 bp in length

* 73307 73406: gap of unknown length

* 73407 77576: contig of 4170 bp in length

* 77577 77676: gap of unknown length

* 77677 79917: contig of 2241 bp in length

* 79918 80017: gap of unknown length

* 80018 83098: contig of 3081 bp in length

83099	83198: gap of unknown length
83199	85948: contig of 2750 bp in length
85949	86048: gap of unknown length
86049	88300: contig of 2252 bp in length
88301	88400: gap of unknown length
88401	90834: contig of 2433 bp in length
90835	90934: gap of unknown length
90935	92486: contig of 1552 bp in length
92487	92586: gap of unknown length
92587	94928: contig of 2342 bp in length
94929	95028: gap of unknown length
95029	97296: contig of 2268 bp in length
97297	97396: gap of unknown length
97397	99357: contig of 1961 bp in length
99358	99457: gap of unknown length
99458	101633: contig of 2176 bp in length
101634	101733: gap of unknown length
101734	103947: contig of 2214 bp in length
103948	104047: gap of unknown length
104048	106350: contig of 2303 bp in length
106351	106450: gap of unknown length
106451	108776: contig of 2326 bp in length
108777	108876: gap of unknown length
108877	111586: contig of 2710 bp in length
111587	111686: gap of unknown length
111687	113940: contig of 2254 bp in length
113941	114040: gap of unknown length
114041	115113: contig of 1073 bp in length
115114	115213: gap of unknown length
115214	116454: contig of 1241 bp in length
116455	116554: gap of unknown length
116555	118145: contig of 1591 bp in length
118146	118245: gap of unknown length
118246	119842: contig of 1557 bp in length
119843	119942: gap of unknown length
119943	121213: contig of 1271 bp in length
121214	121313: gap of unknown length
121314	123957: contig of 2644 bp in length
123958	124057: gap of unknown length
124058	125239: contig of 1162 bp in length
125239	125339: gap of unknown length
125340	126844: contig of 1505 bp in length
126845	126944: gap of unknown length
126945	128704: contig of 1760 bp in length
128705	128804: gap of unknown length
128805	130980: contig of 2176 bp in length
130981	131080: gap of unknown length
131081	132857: contig of 1777 bp in length
132858	132957: gap of unknown length
132958	134901: contig of 1944 bp in length
134902	135001: gap of unknown length
135002	136075: contig of 1074 bp in length
136076	136175: gap of unknown length
136176	137403: contig of 1228 bp in length
137404	137503: gap of unknown length
137504	139120: contig of 1617 bp in length
139121	139220: gap of unknown length
139221	141101: contig of 1861 bp in length
141102	141201: gap of unknown length
141202	143536: contig of 2335 bp in length
143537	143636: gap of unknown length
143637	144839: contig of 1203 bp in length
144840	144939: gap of unknown length
144940	146098: contig of 1159 bp in length
146099	146198: gap of unknown length
146199	147377: contig of 1179 bp in length
147378	147477: gap of unknown length
147478	148718: contig of 1241 bp in length
148719	148818: gap of unknown length

alignment_scores:

Quality: 1395.00

Length: 306

Ratio: 4.697 Gaps: 0

193021 TTTCACATGGTCCCTGGGGGCTGAGATGGTGCCTTTACTACATGGCCTA 192972

Rockville, MD 20830, USA

FEATURES

Location/Qualifiers

source 1. .926
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"

gene
/gene="olfactory receptor MOR251-4P"
/note="potential coding region disrupted by one or more internal stop codons"
/pseudo
/evidence-not-experimental

BASE COUNT 227 a 209 c 181 g 309 t
ORIGIN

alignment_scores:
Quality: 1354.00 Length: 308
Ratio: 4.544 Gaps: 2
Percent Similarity: 96.753 Percent Identity: 84.091

Alignment_block:
us-09-975-308-9 x AY073900 ..

Align seg 1/1 to: AY073900 from: 1 to: 926

```

1 MetAsnHisSerValAlaThrGluPheLeuLeuGlyLeuThrLys 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 ATGAACTTCACGATTGTTAGTAGATTATGATGCTGGAGCTTACTCAAAA 53
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 spGluLeuGlnGlyLeuLeuPheLeuPheLeuLeuValTyrLeu 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 GTGCAACTTCAGGGAATACCTTTCATGTTCTTTTATCTTACCTG 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 AlaAlaPheLeuGlyAsnMetLeuLeuLeuLeuAlaLysLeuSerAsn 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 TGGCTCTACTGTTGTAATGCTAAATGTTGTCATATCTTAACACC 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ThrLeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAla 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 ACCTTGACACACCCATGATATATCTCTTGGCCTTGGCTGTGCGA 203
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 PheLeuGlyThrThrSerLeuLeuProLysMetLeuGlyThrMetLeu 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 CATATCTGCATCAAGCATCATACCCAAATGTTGGAACTAGTTAA 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 hSerGluAsnThrLeuSerTyrAlaGlyCysMetSerGlnLeuPheLeu 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 CATTCAAAAAATTCATTTATATGAGGGGTTGCATGCTCCAGCTCTT 303
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 PheThrTyrSerLeuGlyAlaGluMetValLeuPheThrThrMetAla 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 TTCACATGGTCTCTGGGGCTGAGATGGTCTTACTACAAATGGCTTA 353
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 rAsparGlyTyrValAlaLeuGlyPheProLeuHisTyrSerThrLeu 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 TGACCCGCTATGTCGCACTTGTCCACTTCGCTATAGTACTATTATA 403
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaLeuAla 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 ACCACTATACATGTGATGGCTTGTAGCATGTGCTATGCTATGCTGA 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ThrAsnSerTyrValHisThrAlaLeuLeuMetArgLeuThrPheCys 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
454 ACCAAATTCCTGGTGCACACTGCTCATCTTGAGGCTGACTTCTGTG 503
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLeu 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 GCGCAATATATATGACCACTTCTTCTGTAATATACCCCGCTGCTTG 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 aLeuSerCysSerProValArgIleAsnGluValMetValTyrValAla 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
554 TTCTCTCTGATGCTGTAGAGATCATGAGTATGATGCTATGCTGCTG 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 splLeuThrLeuAlaIleGlyAspPheLeuLeuThrCysIleSerTyr 216

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 ATATACCCCTGGCTGTGGAGACTTACTATACCTGCATTCTCTATGA 653
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 PheLeuIleValAlaIleLeuArgIleArgThrValGluGlyLysArg 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 TTTATATATGCTGCTATCTCTCCGATCCGACACAGAACGACAAAGAA 703
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 sAlaPheSerThrCysSerSerHisLeuThrValValThrLeuTyr 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 GGCCTTCTTACTGCTCTATGCACTGACCTCATGTGTGCTGCTTACTAT 753
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 eProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGlu 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 CTCCTGTATCTACACCTATATCCAGCTGCATCCAGATACCTTGTAT 803
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 ArgAspLysValValAlaAlaLeuTyrThrLeuValThrProThrLeu 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
804 AAGACAAAGGTGTGCTGCTCATATATATCTAGTATACCTTACATGAA 853
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 nProMetValTyrSerPheGlnAsnArgLysMetGlnAlaGlyIleArg 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
854 CCCATATAGTATATAGCTTCAGGAACAAAGATGCTCAGGAATTAAGA 903
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 yValAlaPheAlaPheLeuLys 306
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
904 AAGTATTTGCATTTTAAAA 923
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_pat:AX350633
seq_documentation_block:
LOCUS AX350633 798 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 11 from Patent WO01771177.
ACCESSION AX350633
VERSION AX350633.1 GI:18616201
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites) Gangoli, E.A., Rastelli, L., Smithson, G., Padigaru, M.,
Vernet, C.A., Wolenc, A.R., Casman, S.J., Tcheney, V.T.,
Szekeres, E.S., Gorisse, W., Alsobrook, D.P. and Burgess, C.E.
Novel gPCR-proteins and nucleic acids encoding same
JOURNAL Patent: WO 0177117-A 11 18-OCT-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source 1. .798
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 191 a 202 c 164 g 241 t
ORIGIN
```

alignment_scores:
Quality: 1339.00 Length: 259
Ratio: 5.170 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.228

Alignment_block:
us-09-975-308-9 x AX350633 ..

Align seg 1/1 to: AX350633 from: 1 to: 798

```

49 SerAsnThrLeuHisThrProMetTyrValPheLeuLeuThrLeuAla 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 AACCAACCTTCATACGCGCATGATGTTTCTTCCATGCACTGGCTGT 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 lValAspIleLeuGlyThrThrSerLeuLeuProLysMetLeuGlyThr 82
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 TGTGGACATCATCTGCACAAACGATCATACCGAAGATGCTGGGGACCA 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 eLeuThrSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeu 98

```

```

|||||
103 TGTACATAGAGAAATACATTTCATATGCAGGCTCCAGTCCCACTC 152
99 PheLeuPheThrTrpSerIleuGIyAlaGluMetValIleuPheThrTMe 115
153 TTCTTTGTCACATAGTCTCTGGAGGCTGAGATGTTCTCTTCACACCAT 202
115 TAlaTyRAspArgTrpValAlaIleCysPheProLeuHisTyRSerThrI 132
203 GGCTATGACCGCTATGTGGCATTGTTCCTTCCTTCATACAGTACTG 252
132 LeuEtaSnHisMetCysValAlaIleuSerMetValMetAlaIle 148
253 TTATGACACCATATGTGTGTGCTGCTGCTGACATGGCATGGCATATT 302
149 AlaValThrAsnSerTrpValHisThrAlaIleuIleMetArgLeuThrP 165
303 GCAGTCACCAATTCCTGGGTGCACACAGCTTATCATGAGTTGACTTT 352
165 eCysGlyProAsnThrIleAspHisPheCysGluIleProLeuL 182
353 CTGTGGCCAAACACCATTTGACCATCTTCTGTGAGATACCCCATGTC 402
182 euAlaLeuSerCysSerProValArgIleAsnGluValMetValTyRVal 198
403 TGGCTTTGCTCTGACCCCTGTAAGATCATATGAGGTGATGTTGTT 452
199 AlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrcysIleSerTy 215
453 GCTGATATTACCTGGCCATAGGGGACTTATTTACCTGATCTGCTTA 502
215 rGlyPheIleIleValAlaIleLeuArgIleArgThrValGluGlyAs 232
503 TGTGTTTATCATTTGTGCTATTCTCCGATACCGACAGTAGAAGGACA 552
232 rGlySaIaPheSerThrCysSerSerHisLeuThrValAlaThrLeuTy 248
553 GGAAGGCTCTTCACATGCTCATCTCATCTCAGAGTGAGCCCTTAC 602
249 TyRSerProValIleTyRThrTyRThrIleArgProAlaSerSerTyRTh 265
603 TATTCCTCGTATCTACACCTATATCCGCCCTGCTCCAGCTATACATT 652
265 eGluAlaGAspLysValIleAlaIleLeuTyRThrLeuValThrProTh 282
653 TGAAGAGACAGAGGTGAGTCTCATTCTTGTGACTCCACAT 702
282 euAsnProMetValTyRSerPheGluAsnArgIleMetGluAlaGlyLe 298
703 TAAACCCGATGTGTACAGCTTCAGAAATAGGAGATGACAGGAGAAAT 752
299 ArgLysValPheAlaPheLeuLysHis 307
753 AGGAGGTGTTCATTCTGAAACAC 779

```

seq_name: gb_pat:AX350675

seq_documentation_block:

LOCUS AX350675 663 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 53 from Patent WO0177177.
 ACCESSION AX350675
 VERSION AX350675.1 GI:18616221
 KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Vernet,S., Gangolli,E.A., Rastelli,L., Smithson,G., Padigar,M.,
 Szekeres,E.S., Gorse,W., Alsbrook,J.P., and Burgess,C.E.

TITLE

Novel gpcr-proteins and nucleic acids encoding same
 JOURNAL Patent: WO 0177177-A 53 18-Oct-2001;

Curagen Corporation (US)
 FEATURES Location/Qualifiers
 source 1..663
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 151 a 174 c 134 g 204 t
 ORIGIN

alignment_scores:
 Quality: 1143.00 Length: 221
 Ratio: 5.172 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.548

alignment_block:
 US-09-975-308-9 x AX350675 ..

Align seg 1/1 to: AX350675 from: 1 to: 663

```

63 LeuAlaValAlaAspIleIleCysThrThrSerIleIleProLysMetLe 79
1 CTGGCTGTTGTGACATCATCTGCACACACAGATCATACCGAAGATGCT 50
79 uGlyThrMetLeuThrSerGluAsnThrIleSerTyRAlaGlyCysMet 96
51 GGGGACCATCTCAACATCAGAAATACATTTCATATGCAGGCTCATGT 100
96 eGluLeuPheLeuPheThrTrpSerIleuGIyAlaGluMetValIleuphe 112
101 CCAGGCTCTCTGTTACATGCTCTGTGGAGCTGAGATGTTCTCTTC 150
113 ThrThrMetValTyRAspArgTrpValAlaIleCysPheProLeuHisTy 129
151 ACCACCATGCGCTATGACCGCTATGTGGCCATTGTTCCTCTTCATTA 200
129 rSerThrIleMetAsnHisMetCysValAlaIleLeuLeuSerMetValM 146
201 CACTACTGTTATGACACCATATGTGTGAGCTTGCTGACATGATGTA 250
146 eAlaIleAlaValAlaThrAsnSerTrpValHisThrAlaIleuIleMetArg 162
251 TGGCTATGTGACACCAATTCCTGGGTGCACACACTCTTATCATGAGG 300
163 LeuThrPheCysGlyProAsnThrIleAspHisPhePheCysGluIlePr 179
301 TTGACTTTTCTGTGGCCAAACACCATTTGACCACTTCTTCTGTAGATACC 350
179 oProLeuLeuAlaLeuSerCysSerProValArgIleAsnGluValMetY 196
351 CCCATTGCTGGCTTTGCTGTGAGCCCTGTGAAGATCAATGAGGTGATG 400
196 aTyRValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrcys 212
401 TGTATGTGTGATATTACCTGCGCATAGGGACCTTATTTCTTACTCTG 450
213 IleSerTyRGIyPheIleIleValAlaIleLeuArgIleArgThrValGI 229
451 ATTCTCTATGTTTATCATTTGCTATTTCTCGATATCCGACAGTAGA 500
229 uGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThrValAlaTh 246
501 AGGCAAGAGAGAGCGCTTCACACATGCTCATCTCATCTGACAGTGTGA 550
246 hIleuTyRTrpSerProValIleTyRThrTyRThrIleArgProAlaSer 262
551 CCTTACTATTCTCCTGTAATCTACACCTATATCCGCCCTGCTTCACAC 600
263 TyRThrPheGluArgAspLysValAlaIleAlaIleLeuTyRThrLeuValTh 279
601 TATACATTTGAAAGACAGAGGTGAGTCTGACACTTATATCTTGTGAC 650
279 rProThrLeuAsn 283
|||||

```

651 TCCACATTAAC 663

seq_name: gp_htg:AC101272

seq_documentation_block:

AC101272 59688 bp DNA linear HTG 23-NOV-2001

DEFINITION Mus musculus clone RP23-10J20, LOW-PASS SEQUENCE SAMPLING.

AC101272

AC101272.1 GI:17060047

HTG: HTGS_PHASE0.

KEYWORDS

SOURCE

house mouse

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 59688)

Mus musculus, clone RP23-10J20

Unpublished

2 (bases 1 to 59688)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,

Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,

Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,

Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagoe, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,

Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,

Menus, L., Milova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strassus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16346

Center clone name: 101_J_20

* NOTE: This record contains 76 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely for

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 679: contig of 679 bp in length

* 680 779: gap of 100 bp

* 780 1489: contig of 710 bp in length

* 1490 1589: gap of 100 bp

* 1590 2263: contig of 674 bp in length

* 2264 2363: gap of 100 bp

2364 3054: contig of 691 bp in length

* 3055 3154: gap of 100 bp

* 3155 3839: contig of 685 bp in length

* 3840 3939: gap of 100 bp

* 3940 4630: contig of 691 bp in length

* 4631 4730: gap of 100 bp

* 4731 5401: contig of 671 bp in length

* 5402 5501: gap of 100 bp

* 5502 6164: contig of 663 bp in length

* 6165 6264: gap of 100 bp

* 6265 6949: contig of 665 bp in length

* 6950 7049: gap of 100 bp

* 7050 7728: contig of 679 bp in length

* 7729 7828: gap of 100 bp

* 7829 8516: contig of 688 bp in length

* 8517 8616: gap of 100 bp

* 8617 9325: contig of 709 bp in length

* 9326 9425: gap of 100 bp

* 9426 10127: contig of 702 bp in length

* 10128 10227: gap of 100 bp

* 10228 10924: contig of 697 bp in length

* 10925 11024: gap of 100 bp

* 11025 11712: contig of 688 bp in length

* 11713 11812: gap of 100 bp

* 11813 12481: contig of 669 bp in length

* 12482 12581: gap of 100 bp

* 12582 13254: contig of 673 bp in length

* 13255 13354: gap of 100 bp

* 13355 14033: contig of 679 bp in length

* 14034 14133: gap of 100 bp

* 14134 14816: contig of 683 bp in length

* 14817 14916: gap of 100 bp

* 14917 15597: contig of 681 bp in length

* 15598 15697: gap of 100 bp

* 15698 16388: contig of 691 bp in length

* 16389 16488: gap of 100 bp

* 16489 17234: contig of 746 bp in length

* 17235 17334: gap of 100 bp

* 17335 18009: contig of 675 bp in length

* 18010 18109: gap of 100 bp

* 18110 18802: contig of 693 bp in length

* 18803 18902: gap of 100 bp

* 18903 19608: contig of 706 bp in length

* 19609 19708: gap of 100 bp

* 19709 20385: contig of 677 bp in length

* 20386 20485: gap of 100 bp

* 20486 21178: contig of 693 bp in length

* 21179 21278: gap of 100 bp

* 21279 21942: contig of 664 bp in length

* 21943 22042: gap of 100 bp

* 22043 22714: contig of 672 bp in length

* 22715 22814: gap of 100 bp

* 22815 23486: contig of 672 bp in length

* 23487 23586: gap of 100 bp

* 23587 24261: contig of 675 bp in length

* 24262 24361: gap of 100 bp

* 24362 25046: contig of 685 bp in length

* 25047 25146: gap of 100 bp

* 25147 25807: contig of 661 bp in length

* 25808 25907: gap of 100 bp

* 25908 26596: contig of 689 bp in length

* 26597 26696: gap of 100 bp

* 26697 27406: contig of 710 bp in length

* 27407 27506: gap of 100 bp

* 27507 28223: contig of 717 bp in length

* 28224 28323: gap of 100 bp

* 28324 29026: contig of 703 bp in length

* 29027 29126: gap of 100 bp

* 29127 29814: contig of 688 bp in length

* 29815 29914: gap of 100 bp

* 29915 30588: contig of 674 bp in length

* 30589 30688: gap of 100 bp

* 30689 31361: contig of 673 bp in length

[illegible]

Assignment_2.docx	Quality:	964.00	Length:	213
	Ratio:	4.657	Gaps:	0
Percent Similarity:		97.183	Percent Identity:	84.977

```
alignment_block:
US-09-975-308-9 x AC101272
```

Align seg 1/1 to: AC101272 from: 1 to: 59688

14 LeuThrLysLysProGluLeuGlnCylLeilepheLeuPheLeuI 30
|||||:||||| |||||||||:|||||: |||||:
25170 CTTACTCAAAAGCTGGAACCTTCAGGSAATACTTTTCATGTCTTCTTTT 25219

[illegible]

[illegible]

```

498 CTCCAAATGCGTTGAGCAGCTTCTCTGTGGAGATTCGCCCACTGTTGAAGC 547
184 euserCysSerProValArgIleasnIuValMetValTyrValAlaAsp 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
548 TCTCTGTGCTCCAAACAATTTGAACAGGCGCATGGCTTGTGCTGCAGAT 597
201 IleThrIleuAlaIleGlyAspPheIleuThrCysIleSerTyrGlyPhe 217
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
598 GTGTTCCTCGGCTGTAGGGAAGCTTCTCTGTGATCATCTCTCTCTAGGCTT 647
217 eIleIleValAlaIleuArgIleArgTyrValIleuGlyAspGlyAsp 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
648 TATGTGGCCAGCACTCGAAGATCCGCTCAGCTGAGGGCAAGCAACGAG 697
234 IapheSerThrCysSerSerHisIleuThrValIleThrLeuTyrTyrSer 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
698 CCTTCTCCACCTGCTCTGCACACCTCATCGTCGTGTGTACCATGACTACTCC 747
251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArg 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
748 ACTGCATCATACACTACATCTGCGCTTCATCCACTCAGCTACTCAACAA 797
267 GASPLYSValValAlaAlaLeuTyrThrLeuValThrProThrIleAsp 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
798 GGACAGAGTGGTGTCATCATCTACACCTCATGTCAGGCCACCCACTTGAAAC 847
284 rometValTyrSerPheGlnAsnArgIleuMetGlnAlaGlyIleArgIys 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
848 CTCCTATCTACACTCTGAGGAGTAAGATGTCAAGTTGCATCGGAGAG 897
301 ValPheAla 303
:::|||||:
898 CTTTCTCTCC 906

seq_name: gb AY074180

seq_documentation_block:
LOCUS AY074180 916 bp DNA linear ROD 04-FEB-2002
DEFINITION Mus musculus olfactory receptor MOR251-5 pseudogene, partial
sequence.
ACCESSION AY074180
VERSION AY074180.1 GI:18481342
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 916)
Zhang, X. and Firestein, S.
The olfactory receptor gene superfamily of the mouse
Nat. Neurosci. 5 (2), 124-133 (2002)
11802173
2 (bases 1 to 916)
Adams, M.
Direct Submission
Submitted (11-JAN-2002) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
FEATURES
source
1..916
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="7"
<1..>916
/gene="Olfactory receptor MOR251-5"
/note="potential coding region disrupted by one or more
internal stop codons"
/pseudo
/evidence=not_experimental
BASE COUNT 185 a 287 c 200 g 244 t
ORIGIN
alignment_scores:

```

Quality: 938.50 Length: 304
 Ratio: 3.450 Gaps: 2
 Percent Similarity: 89.474 Percent Identity: 55.592

alignment block:
 US-09-975-308-9 x AY074180 ..

Align seg 1/1 to: AY074180 from: 1 to: 916

```

1 MetasHisSerValValThrGluPheIleIleuGlyLeuThrIleuSly 17
  ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 ATGATGAGACACTGCTACTGAGTCTCTCATCTCGGATTCAGAAAT 53
17 sProGluLeuGlnGlyIleIlePheLeuPhePheIleValIleuThr 34
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 GCCTCACCTTCGGTACCTCTTCTTCTCAGCTTCTCTGCTCATACATG 103
34 aAlaPheLeuGlyAsnMetLeuIleIleIleAlaIleuIleuSerAsn 50
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 CTGCAATCTCAGAACTGCTCATATGATGATGATGATGATGATGATGAT 153
51 ThrLeuHisThrProMetIleValPheLeuLeuThrIleuAlaValAs 67
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 GCCCTGCATACCCCATGCTCTCTCTGCTCAACTGGCCATCTTGGA 203
67 PileIleCysThrThrSerIleIleProIleuMetLeuGlyThrMetLeu 84
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 CATCTCTGCATCTCCACCATCTCCACCATCTCTGATGATGATGATGAT 253
84 hSerGluAsnThrIleSerIleValIleGlyCysMetSerGlnLeuPhe 100
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 CAGGA...AGGACCATCTCTACGGGGGCGCATGGCCAGCTCTCTTC 300
101 PheThrIlePheSerIleuGlyAlaGluMetValLeuPheThrIleuAla 117
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TTACATGATGCTACCTGGAGAGACACTCTCTCTCTCTCTCTCTCTCT 350
117 rAspArgTyrValAlaIleCysPheProLeuHisThrIleMetAla 134
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 TGACCGCTTGTGGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
134 snHisIleMetCysValAlaIleuSerMetValMetAlaIleAlaVal 150
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 GCCCGCAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450
151 ThrAsnSerTyrValHisThrAlaLeuIleMetArgLeuThrPheCysG 167
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 ACTAACACACGCTGACACAGCGCTGATGCTGCTGCTGCTGCTGCTG 500
167 yProAsnThrIleAspHisPhePheCysGluIleProLeuLeuAla 184
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 CTCGCAATGTATAGAGCATCTTCTGTGAGATTCCTCCCTGTTGAGC 550
184 euserCysSerProValArgIleAsnGluValMetValTyrValAlaAsp 200
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 TTTCCTGCTGCTCAAGCATGTGATGAGGCAATGGCTGCTGCTGCTG 600
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGly 217
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 GTTCTCTGCTGCTGAGGAACTTCTGTGCAATCTCTCTCTCTCTCT 650
217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyArgIle 234
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 TATTGAGCCAGCATCTGAGATCCGCTGAGCGAGGAGGAGGAGGAGG 700
234 laphSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSer 250
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 CCTTCTCAGCTGCTGCTGCAACCTATGCTGCTGCTGCTGCTGCTG 750
251 ProValIleTyrThrIleArgProAlaSerSerTyrThrPheGluArg 267
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 ACTGTCTATCTACACTACATTCGCCCTCTCTCTCTCTCTCTCTCTCT 800

```

```

267 gAspLysValValAlaLeuThrThrIleuValThrProThrLeuAsn 283
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 GGACAAAGTGGGTGCTCATCTCTACACCTCAGTGGACCCACTTGAAC 850
284 ProMetValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArg 300
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 CCCCTCATCTACACCTGAGGAGCAAGATGTCAAGTGTGCTGCTGAG 900
300 sValPheAla 303
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 ACTTCTCTCC 910
seq_name: gb_hhg:AC094870
seq_documentation_block:
LOCUS AC094870 112730 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-5N18. *** SEQUENCING IN PROGRESS ***
ACCESSION AC094870
VERSION AC094870.2 GI:17941654
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 112730)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blinige,K., Blankenburg,K., Bonini,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferriguto,D., Flagg,N., Ford,J.,
Foster,P., Frintz,P., Gabist,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
Hollins,B., Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvath,J.,
Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mayhew,B., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,D., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S.,
Ogub,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prins,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,A., Shooster,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 112730)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624706.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBPG

Center clone name: CH230-5N18

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findphraplist

Consensus quality: 96697 bases at least Q40

Consensus quality: 104847 bases at least Q30

Consensus quality: 112341 bases at least Q20

Estimated insert size: 92352; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

(see <http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html>).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 52 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```

1      6916: contig of 6916 bp in length
*      6917      7016: gap of unknown length
*      7017      11089: contig of 4073 bp in length
*      11090      11189: gap of unknown length
*      11190      14517: contig of 3328 bp in length
*      14518      14617: gap of unknown length
*      14618      17992: contig of 3375 bp in length
*      17993      18092: gap of unknown length
*      18093      21658: contig of 3566 bp in length
*      21659      21758: gap of unknown length
*      21759      25307: contig of 3549 bp in length
*      25308      25407: gap of unknown length
*      25408      27818: contig of 2411 bp in length
*      27819      27918: gap of unknown length
*      27919      31526: contig of 3608 bp in length
*      31527      31626: gap of unknown length
*      31627      33967: contig of 2341 bp in length
*      33968      34067: gap of unknown length
*      34068      36642: contig of 2575 bp in length
*      36643      36742: gap of unknown length
*      36743      39205: contig of 2463 bp in length
*      39206      39305: gap of unknown length
*      39306      41281: contig of 1976 bp in length
*      41282      41381: gap of unknown length
*      41382      45065: contig of 3684 bp in length
*      45066      45165: gap of unknown length
*      45166      47143: contig of 1978 bp in length
*      47144      47243: gap of unknown length
*      47244      48499: contig of 1256 bp in length
*      48499      48599: gap of unknown length
*      48599      50055: contig of 1456 bp in length
*      50055      50155: gap of unknown length
*      50156      53530: contig of 3375 bp in length
*      53531      53630: gap of unknown length
*      53631      55445: contig of 1815 bp in length
*      55446      55545: gap of unknown length
*      55546      58015: contig of 2470 bp in length
*      58015      58115: gap of unknown length
*      58116      59612: contig of 1497 bp in length
*      59613      59712: gap of unknown length
*      59713      61702: contig of 1990 bp in length
*      61703      61802: gap of unknown length
*      61803      63157: contig of 1355 bp in length

```

```

*      63158      63257: gap of unknown length
*      63258      64623: contig of 1366 bp in length
*      64624      64723: gap of unknown length
*      64724      65766: contig of 1043 bp in length
*      65767      65866: gap of unknown length
*      65867      67869: contig of 2003 bp in length
*      67870      67969: gap of unknown length
*      67970      69560: contig of 1591 bp in length
*      69561      69660: gap of unknown length
*      69661      71980: contig of 2320 bp in length
*      71981      72080: gap of unknown length
*      72081      73657: contig of 1577 bp in length
*      73658      73757: gap of unknown length
*      73758      75695: contig of 1938 bp in length
*      75696      75795: gap of unknown length
*      75796      78395: contig of 2600 bp in length
*      78396      78495: gap of unknown length
*      78496      79707: contig of 1212 bp in length
*      79708      79807: gap of unknown length
*      79808      81374: contig of 1567 bp in length
*      81375      81474: gap of unknown length
*      81475      83163: contig of 1689 bp in length
*      83164      83263: gap of unknown length
*      83264      85474: contig of 2211 bp in length
*      85475      85574: gap of unknown length
*      85575      87264: contig of 1690 bp in length
*      87265      87364: gap of unknown length
*      87365      88370: contig of 1006 bp in length
*      88371      88470: gap of unknown length
*      88471      89719: contig of 1249 bp in length
*      89720      89819: gap of unknown length
*      89820      91531: contig of 1712 bp in length
*      91532      91631: gap of unknown length
*      91632      92853: contig of 1222 bp in length
*      92854      92953: gap of unknown length
*      92954      95061: contig of 2108 bp in length
*      95062      95161: gap of unknown length
*      95162      97151: contig of 1990 bp in length
*      97152      97251: gap of unknown length
*      97252      98827: contig of 1576 bp in length
*      98828      98927: gap of unknown length
*      98929      100300: contig of 1733 bp in length
*      100301      100400: gap of unknown length
*      100401      101467: contig of 1067 bp in length
*      101468      101567: gap of unknown length
*      101568      102772: contig of 1205 bp in length
*      102773      102872: gap of unknown length
*      102873      104128: contig of 1236 bp in length
*      104129      104228: gap of unknown length
*      104230      10514: contig of 1286 bp in length
*      10515      105615: gap of unknown length
*      105616      106797: contig of 1183 bp in length
*      106798      106897: gap of unknown length
*      106898      108212: contig of 1315 bp in length
*      108213      108313: gap of unknown length
*      108313      109627: contig of 1315 bp in length
*      109628      109727: gap of unknown length
*      109728      110921: contig of 1194 bp in length
*      110922      111022: gap of unknown length
*      111022      112730: contig of 1709 bp in length.

```

FEATURES

source 1.112730
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-5N18"

BASE COUNT 30919 a 22306 c 22742 g 31565 t 5198 others
ORIGIN

alignment_scores: 935.00 Length: 303
Quality: 3.476 Gaps: 2
Ratio: 88.779 Percent Identity: 55.446
Percent Similarity: 88.779

Align seg 1/1 to reverse of: AC094870 from: 1 to: 112730

REFERENCE	AUTHORS	seq. documentation_block:	197438 bp	DNA	linear	HTG 20-DEC-2001
284	romerallyterPheginsnasnglucMerclnlaaglyleatyls	300				
95497	CCCTATCTACACCTGAGGAACAAGATGTCAAACTTCACCTCGGAAA	95448				
301	Valpneala 303					
95447	CTTCGTCC 95439					
seq_name:	gb.htg:AC094700					
LOCUS	AC094700	197438 bp	DNA	linear	HTG 20-DEC-2001	
DEFINITION	Rattus norvegicus clone CH230-5F7, *** SEQUENCING IN PROGRESS ***					
ACCESSION	AC094700					
VERSION	AC094700.4	GI:17941479				
KEYWORDS	HTG; HTGS; PHASE1.					
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
REFERENCE	1. (bases 1 to 197438)					
AUTHORS	Mizny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Ameratunge,H.C., Are,J.R., Banks,T., Barbara,J., Beaton,J., Bilmage,K., Blankenburg,K., Bonni,D., Bouck,J., Brown,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carrott,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Demu,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunnarne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,D., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louseghd,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,K., Mel,G., Melker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkw,S., Ogun,M., Okunolu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quills,M., Ren,Y., Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodegren,E., Sonike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellro,B., Thoms,N., Thoms,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.					
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 197438)					
AUTHORS	Worley,K.C.					
TITLE	Direct Submission					
JOURNAL	Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
COMMENT	On Dec 20, 2001 this sequence version replaced g1:17062158.					
	Genome Center					

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: G8W
Center clone name: CH230-5F7
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329first call to findhaplist
Consensus quality: 163315 bases at least Q40
Consensus quality: 177813 bases at least Q30
Consensus quality: 189359 bases at least Q20
Estimated insert size: 163704; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 96 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6381 6380: contig of 6380 bp in length
6481 6480: gap of unknown length
11235 11235: contig of 4755 bp in length
11335 11335: gap of unknown length
14082 14082: contig of 2747 bp in length
14083 14182: gap of unknown length
14183 18911: contig of 4729 bp in length
18911 22084: gap of unknown length
19012 22084: contig of 3073 bp in length
22085 22184: gap of unknown length
22185 26199: contig of 4015 bp in length
26200 26299: gap of unknown length
26300 28568: contig of 2269 bp in length
28569 28668: gap of unknown length
28669 31639: contig of 2971 bp in length
31640 31739: gap of unknown length
31740 33776: contig of 2037 bp in length
33777 33876: gap of unknown length
33877 37642: contig of 3766 bp in length
37643 37742: gap of unknown length
37743 39937: contig of 2195 bp in length
39938 40037: gap of unknown length
40038 43330: contig of 3293 bp in length
43331 43430: gap of unknown length
43431 46413: contig of 2983 bp in length
46414 46513: gap of unknown length
46514 49924: contig of 3411 bp in length
49925 50024: gap of unknown length
50025 52777: contig of 2753 bp in length
52778 52877: gap of unknown length
52879 55322: contig of 2445 bp in length
55323 55422: gap of unknown length
55423 57118: contig of 1696 bp in length
57119 57218: gap of unknown length
57219 59661: contig of 2443 bp in length
59662 59761: gap of unknown length
59762 62279: contig of 2518 bp in length
62280 62379: gap of unknown length
62380 65763: contig of 3384 bp in length
65764 65863: gap of unknown length
65864 68840: contig of 2977 bp in length
68841 68940: gap of unknown length
68941 70763: contig of 1823 bp in length
70764 70863: gap of unknown length
70864 73785: contig of 2922 bp in length
73786 73885: gap of unknown length

76368: contig of 2483 bp in length
76369 76468: gap of unknown length
76469 79049: contig of 2581 bp in length
79050 79148: gap of unknown length
79150 82088: contig of 2939 bp in length
82089 82188: gap of unknown length
82189 84182: contig of 1994 bp in length
84183 84283: gap of unknown length
84284 87125: contig of 2843 bp in length
87126 87225: gap of unknown length
87226 88998: contig of 1773 bp in length
88999 89098: gap of unknown length
89099 92015: contig of 2917 bp in length
92016 92115: gap of unknown length
92116 94512: contig of 2397 bp in length
94513 94612: gap of unknown length
94613 96876: contig of 2264 bp in length
96877 96976: gap of unknown length
96977 98801: contig of 1825 bp in length
98802 98901: gap of unknown length
98902 100728: contig of 1827 bp in length
100729 100828: gap of unknown length
100829 102598: contig of 1770 bp in length
102599 102698: gap of unknown length
102699 104252: contig of 1554 bp in length
104253 104352: gap of unknown length
104353 105818: contig of 1466 bp in length
105819 105918: gap of unknown length
105919 107016: contig of 1098 bp in length
107017 107116: gap of unknown length
107117 108700: contig of 1584 bp in length
108701 108800: gap of unknown length
108801 110130: contig of 1330 bp in length
110131 110230: gap of unknown length
110231 111400: contig of 1170 bp in length
111401 111500: gap of unknown length
111501 113819: contig of 2319 bp in length
113820 113919: gap of unknown length
113920 115967: contig of 2048 bp in length
115968 116067: gap of unknown length
116068 117164: contig of 1097 bp in length
117165 117264: gap of unknown length
117265 118813: contig of 1549 bp in length
118814 118913: gap of unknown length
118914 120584: contig of 1670 bp in length
120585 120683: gap of unknown length
120684 122741: contig of 2058 bp in length
122742 122841: gap of unknown length
122842 125104: contig of 2263 bp in length
125105 125204: gap of unknown length
125205 127018: contig of 1814 bp in length
127019 127118: gap of unknown length
127119 128590: contig of 1472 bp in length
128591 128690: gap of unknown length
128691 129819: contig of 1129 bp in length
129820 129919: gap of unknown length
129919 131056: contig of 1137 bp in length
131057 131057: gap of unknown length
131058 131156: gap of unknown length
131157 133124: contig of 1968 bp in length
133125 133224: gap of unknown length
133225 135098: contig of 1874 bp in length
135099 135198: gap of unknown length
135199 136613: contig of 1415 bp in length
136614 136713: gap of unknown length
136714 138549: contig of 1836 bp in length
138550 138649: gap of unknown length

alignment_scores: 935.00 Length: 303
Quality: 3.476 Gaps: 2
Ratio: 88.779 Percent Identity: 55.446
Percent Similarity: 88.779
alignment_block:

US-09-975-308-9 x AC094700/rev ..

Align seg 1/1 to reverse of: AC094700 from: 1 to: 197438

```

1 MetasnhiservalValthrighpellelleleuglyleuthrlysty 17
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96743 ATGATAGGACACTGGTCACTGAGTTCCTCATCTGGGATTCAGAAAT 96694

17 sprogileugllyllelepheleupheleullevaltyrleuv 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96693 GCGCTGATACCTCGG...GTACCACTTCTCAGCTTCCTGCTATACATGG 96647

34 alalapheluglyasmetleullellellealalytletyrserasn 50
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96646 CTGCATATTCAGAACTGCTCATATGATGAGTCACTGCGACGCCA 96597

51 thrlenuhistrprometyrvalpheluleuthrleualavalas 67
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96596 GCGCTGATACCTCGGATGATCTCTCTGCTGCACTTACCTGCTGCGA 96547

67 pilellecystrhrserllelelepolysmetleuglythmetleut 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96546 CATCCCTCTGACCTCCACATCCCTACCAAGCTCTGACATCATGTA. 96498

84 hrsergluasnthrileserytralaaglycysmetsergluapheleu 100
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96497 .GGGGGAAGACCATCTCTTATGCTGCTGCATGGCCCGCTCTTCTTC 96450

101 phehrtrpserleuglyalaglumetvalleuphethrthrmelaty 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96449 TTCACATGCGCATGGGGGAGAGCTTCTGCTCTCACACTATGGCTTA 96400

117 rasparaglyrvalalalecysphreproleuhsityrserthrilemea 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96399 TGACCTCTGCTGCGCAATCGCTGCTCCCTCATATGACCTGATGG 96350

134 snishismetcyvalalaleuleusermetvalmetalailealal 150
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96349 GCGCCAGGGGTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 96300

151 thrasnsertrpvalhistrhralaleulemetargyleuthrphcysgl 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96299 ACTAACAACAGAGTGTGAACACACCTGCTGCTCATCTACATCTGCAG 96250

167 yproasnthrileasphisphehcygsluileproleuleleual 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96249 CTCATATGATAGACACTCTCTGCTGATGATCCCGCTGTAAAGC 96200

184 eusercysserprovalargileasngluvalmetvaltyrvalalasp 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96199 TCTCTGTGCTCCACACAAATGACAGAGTTATGGCTTGTGTCAGAT 96150

201 llerhthleualalecylaspheulleuthrhcysileseryrlyph 217
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96149 GTCCTTCTGCTGCTAGGAACCTCTGCTGATCATCTCTCTCATGCTT 96100

217 eillellevalalaleuargylearghrvalgluuglylsarglysa 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96099 TATGTGACAGATATCTGATAGATGCTCACTGAGGGGCAACGAG 96050

234 lapheserthrcysserserhisleuthrvalalvalthryrtyrser 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96049 CCGTTTCTACCTGTCTGCAATCTCTTGTGTCACATGATCTACTCC 96000

251 provalilletrthryrleargproalaseretyrthrphelguar 267
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95999 ACTATCATCTACACTACATTCGCTGATCCAGCTACATCATTAACAA 95950

267 gasplysvalalalaleuargylearghrvalgluuglylsarglysa 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95949 GGCACAGGGGTGCTCATCTATCACTGCTGAGGCAACCCACTTGAACC 95900

284 rometvaltyrserphegluasnarglumetlinalaglylearglys 300

```

95899 CCGCTATCTACACTCTGAGGACAAAGATGTCAAGTTCAGCTCGGAAA 95850

301 Valphea1a 303

95849 CTCTGTGCC 95841

seq_name: gb_htg:AC096461

seq_documentation_block:

LOCUS AC096461

DEFINITION Rattus norvegicus clone CH230-58u17, *** SEQUENCING IN PROGRESS

ACCESSION AC096461.2 GI:17955456

VERSION HTG: HTGS_PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 204143)

Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarta,J.,

Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burich,P., Burrell,C., Burrell,K.L., Byrd,N.C., Catron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davys,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

Hollins,B., Homsli,F., Howard,S., Huber,A., Hulky,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Koryah,J.,

Kovar,C., Kratovic,D., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louiseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Lune,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFSC
Center clone name: CH230-58J17
----- Summary Statistics
Assembly Program: Phrap; version 0.990329First call to
findPhrapblast
Consensus quality: 156802 bases at least Q40
Consensus quality: 170292 bases at least Q20
Consensus quality: 181161 bases at least Q20
Estimated insert size: 177160; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 62 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 9619: contig of 9619 bp in length
9620 9719: gap of unknown length
9720 18930: contig of 9211 bp in length
18931 19030: gap of unknown length
19031 27646: contig of 8616 bp in length
27647 27746: gap of unknown length
27747 33370: contig of 5624 bp in length
33371 33470: gap of unknown length
33471 41116: contig of 7646 bp in length
41117 41216: gap of unknown length
41217 48276: contig of 7060 bp in length
48277 48376: gap of unknown length
48377 54436: contig of 6060 bp in length
54437 54536: gap of unknown length
54537 61966: contig of 7430 bp in length
61967 62066: gap of unknown length
62067 66405: contig of 4339 bp in length
66406 66505: gap of unknown length
66506 73151: contig of 6646 bp in length
73152 73251: gap of unknown length
73252 79129: contig of 5878 bp in length
79130 79229: gap of unknown length
79230 85275: contig of 6046 bp in length
85276 85375: gap of unknown length
85376 90469: contig of 5094 bp in length
90470 90569: gap of unknown length
90570 94546: contig of 3977 bp in length
94547 94646: gap of unknown length
94647 99437: contig of 4791 bp in length
99438 99537: gap of unknown length
99538 102811: contig of 3274 bp in length
102812 102911: gap of unknown length
102912 106305: contig of 3394 bp in length
106306 106405: gap of unknown length
106406 111977: contig of 5572 bp in length
111978 112077: gap of unknown length
112079 116021: contig of 3944 bp in length
116022 116121: gap of unknown length
116122 118407: contig of 2286 bp in length
118408 118507: gap of unknown length
118508 121535: contig of 3028 bp in length
121536 121635: gap of unknown length
121636 125077: contig of 3442 bp in length
125078 125177: gap of unknown length
125178 129179: contig of 4002 bp in length
129180 132170: gap of unknown length
132171 132270: contig of 2891 bp in length
132270: gap of unknown length

132271 134079: contig of 1809 bp in length
134080 134179: gap of unknown length
134180 137358: contig of 3179 bp in length
137359 137458: gap of unknown length
137459 139978: contig of 2520 bp in length
139979 140078: gap of unknown length
140079 142972: contig of 2894 bp in length
142973 143072: gap of unknown length
143073 145874: contig of 2802 bp in length
145875 145974: gap of unknown length
145975 147668: contig of 1794 bp in length
147669 147868: gap of unknown length
147869 149593: contig of 1725 bp in length
149594 149693: gap of unknown length
149694 151287: contig of 1594 bp in length
151288 151387: gap of unknown length
151388 153321: contig of 1934 bp in length
153322 153421: gap of unknown length
153422 155471: contig of 2050 bp in length
155472 155571: gap of unknown length
155573 157991: contig of 2420 bp in length
157992 158091: gap of unknown length
158092 160551: contig of 2460 bp in length
160552 160651: gap of unknown length
160652 162741: contig of 2090 bp in length
162742 162841: gap of unknown length
162842 164428: contig of 1587 bp in length
164429 164528: gap of unknown length
164529 166471: contig of 1943 bp in length
166472 166571: gap of unknown length
166572 168429: contig of 1858 bp in length
168430 168529: gap of unknown length
168530 170496: contig of 1967 bp in length
170497 170596: gap of unknown length
170597 172644: contig of 2048 bp in length
172645 172744: gap of unknown length
172745 174365: contig of 1621 bp in length
174366 174465: gap of unknown length
174466 175806: contig of 1341 bp in length
175807 175906: gap of unknown length
175907 178082: contig of 2176 bp in length
178083 178182: gap of unknown length
180094 180094: contig of 1912 bp in length
180095 180194: gap of unknown length
180195 181465: contig of 1271 bp in length
181466 181565: gap of unknown length
181566 183890: contig of 2325 bp in length
183891 183990: gap of unknown length
183991 185397: contig of 1407 bp in length
185398 185497: gap of unknown length
185498 187581: contig of 2084 bp in length
187582 187681: gap of unknown length
187682 188760: contig of 1079 bp in length
188761 188860: gap of unknown length
188861 190644: contig of 1784 bp in length
190645 190744: gap of unknown length
190745 191885: contig of 1141 bp in length
191886 191985: gap of unknown length
191986 193130: contig of 1144 bp in length
193131 193229: gap of unknown length
193230 194840: contig of 1611 bp in length
194841 194940: gap of unknown length
194941 196463: contig of 1523 bp in length
196463: gap of unknown length

alignment_scores:

Quality: 906.00 Length: 306
Ratio: 3.485 Gaps: 0
Percent Similarity: 84.967 Percent Identity: 52.941

alignment_block:

US-09-975-308-9 x AC096461 ..

OM of: US-09-975-308-9 to: N_Geneseq_032802:* out_format : pfs

Date: Jun 24, 2002 12:59 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet_p2n_model -DEV=xlh
-O=/cgn2_1/USPRO/spool/US09975308/runat_24062002_090304_26270/app_query.fasta.1.368
-DB=N_Geneseq_032802 -QEXT=fastcap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPO=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFORMAT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US09975308 -CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-975-308-9
Query length: 307
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 209.350000
```

score_list:

```
Sequence Strd Orig Zscore EScore len 1 Documentation
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32037 + 1574.00 2913.00 6.6e-154
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542353 + 1574.00 2912.96 6.6e-154
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA517179 + 1339.00 2476.78 1.3e-129
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542284 + 861.00 1584.17 6.8e-80
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA12952 + 861.00 1581.75 9.3e-80
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31706 + 860.00 1583.00 7.9e-80
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542455 + 731.00 1342.38 2.0e-66
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542455 + 728.00 1332.76 2.5e-65
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH61193 + 725.00 1331.47 8.1e-66
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32092 + 723.00 1327.44 1.4e-65
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32067 + 707.00 1297.98 6.0e-64
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542407 + 707.00 1296.63 7.1e-64
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32033 + 706.00 1296.12 7.6e-64
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542313 + 706.00 1296.08 7.6e-64
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542310 + 699.00 1282.84 4.2e-63
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542390 + 699.00 1282.54 4.3e-63
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31761 + 697.00 1282.80 4.2e-63
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31886 + 695.00 1275.52 1.1e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542439 + 693.50 1272.40 1.6e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32097 + 693.00 1271.76 1.7e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31177 + 692.00 1273.48 1.4e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32014 + 691.00 1268.06 2.8e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31761 + 691.00 1267.97 2.8e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH33815 + 691.00 1214.83 2.0e-59
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32035 + 687.00 1260.22 7.6e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542360 + 687.00 1259.41 8.4e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31910 + 687.00 1259.42 8.4e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32144 + 686.50 1258.93 8.9e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31717 + 686.00 1262.31 5.8e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31187 + 686.00 1258.59 9.3e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32010 + 686.00 1258.59 9.3e-62
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542357 + 685.00 1253.02 1.9e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH78085 + 685.00 1253.02 1.9e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32420 + 683.50 1254.21 1.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32387 + 683.50 1253.94 1.7e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABK16617 + 683.50 1253.51 1.8e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH31037 + 683.50 1253.10 2.3e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH44109 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH54562 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH24345 + 683.50 1246.17 4.6e-61
```

```
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK02855 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAK8297 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI12861 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI34218 + 683.50 1246.17 4.6e-61
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI02777 + 683.50 1246.17 4.6e-61
```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32037

seq_documentation_block:

ID AAH32037 standard; DNA: 921 BP.

AAH32037;

30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 610.

Human: olfactory receptor; OR; primary scent determination;

secondary scent determination; polypeptide library; odour receptor;

scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

WO200127158-A2.

19-APR-2001.

06-OCT-2000; 2000MO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGT-) DIGSCENTS.

(YEDA) YEDA RES & DEV CO LTD.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I,

WPI; 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 425; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors, and can be used for determining differences in the olfactory faculties of different individuals.

Sequence 921 BP; 218 A; 235 C; 183 G; 285 T; 0 other;

alignment_scores:

Quality: 1574.00 Length: 307
Ratio: 5.127 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.349

alignment_block:

US-09-975-308-9 x AAH32037 ..

Align seg 1/1 to: AAH32037 from: 1 to: 921

1 MetashttservValtrhglupherleileucelyleuthrlyslv 17

```

|||||
1 ATGAATCAGAGCGTTGTAAGTTCATATCTGGGCTTACCAAAA 50
17 sprogileunginglyllelepheleuphepheullevaltyrleu 34
|||||
51 GCCTGAACCCAGGAAATATCTCTCTTTTTCATGTCATCTG 100
34 alalaphelenglyasmetleullellelalyserasn 50
|||||
101 TGCTTTTCGCGCAACATGCTCATCTTCCTCAAAATCTATCAAC 150
51 ThrLeuHisThrPrometyValpheleuThrleuAlaValas 67
|||||
151 ACCTTGATACGCCCATGTATGTTTCTTCTGACACTGGCTGTGGA 200
67 pillelecystrhthserlleleprolysmetleuglythmetleu 84
|||||
201 CATCATGTGACACACAGCATCATCCAGAGATGCTGGGACCATGCTA 250
84 hrsergluasnthrilesertyalaglycysmetserginleupheleu 100
|||||
251 CATGAGAAATACCATTCATATCAGCGTCGATCCAGCTTCTTG 300
101 phenrttrpserleuglyalaglumetvalleupherthrmetalaTy 117
|||||
301 TTCACATGCTCTGGAGCTGAGATGGTCTCTTCACACCATAGGCTA 350
117 rAspArgtyValAlailecyspheproleuHisTySerThrileueta 134
|||||
351 TGACCGCATGTGGCATTTGTTCCCTTCATTACAGTACTGTATGA 400
134 snHisMetCysValAlaileuSerMetValMetAlaileAlaVal 150
|||||
401 ACCACCATATGTGTAGCTTGTCTGACACATGTCATGGCTATTCAGTC 450
151 ThrAsnserTPValHisThrAlaileuilemetArgleuthrphCysgl 167
451 ACCATTTCTGGGTGCACACAGCTTATCATAGAGTTCATCTCTGCG 500
167 yProAsnThrIleasphIsPhephCysGluileProProleuAlaL 184
|||||
501 GCCAACACCATATGACACCTTCTGTGTAGTACCCCATTTGCGGCTT 550
184 euSerCysSerProValArgIleasnGluValMetValTyValAlaasp 200
551 TGTCTGTAGCCCTGTAGAAATCAATGAGGTGATGTGTCTGAT 600
201 llethrleuAlaileglyAspPheileleuthrCysIleSerTyGlyPh 217
|||||
601 ATTACCTTGCCATAGGGGACCTTATCTTACCTGATCTCTATGGTTT 650
217 etileleValAlaileleuArgIleArgThrValGluGlyAsArgLysA 234
|||||
651 TATCATTTGTCATCTCTCGTATCGGACAGTAGAGAGCAAGAGAAAG 700
234 laPheSerThrcysSerSerHisleuthValValThrleuTyTrsSer 250
701 CCTTCAACATGCTCATCTCATCTCACAGTGTGACCCCTTACTATTTCT 750
251 ProValIleTyThrTyrlleArgProAlaSerSerTyThrPheGluAr 267
751 CCTGTATCTACACCTATATCCCGCTGCTCCAGCTATATCAATTGAAG 800
267 gaSPlyValAlaAlaAlaLeuTyThrleuValThrProThrLeuAsn 284
801 AGACAGAGGTGTAGCTGCACTATATCTTGTGACTCCACATTAAC 850
284 roveValTyTrsPheGlnAsnArgLumetGlnAlaGlyIleArgLys 300
851 CGATGTGTACAGCTTCCAGAAATAGGAGATGCGAGAGATTTAGGAG 900
301 ValPheAlaPheLeuLysHis 307
|||||

```

901 GTGTTGCATTTCGAACAC 921

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA542323

seq_documentation_block:

ID AA542323 standard; cDNA: 924 BP.

XX AA542323;

XX 18-DEC-2001 (first entry)

XX Human cDNA encoding olfactory receptor AOLFRL24.

XX Human: olfactory receptor; G protein-coupled receptor; GPCR; odourant; ss; food additive; cosmetic; fragrance; pharmaceutical additive.

XX Homo sapiens.

XX WO200168805-A2.

XX 20-SEP-2001.

XX 13-MAR-2001; 2001WO-US07771.

XX 13-MAR-2000; 2000US-0188914.

XX 24-MAR-2000; 2000US-0192033.

XX 12-APR-2000; 2000US-0198474.

XX 24-APR-2000; 2000US-0199335.

XX 26-MAY-2000; 2000US-0207702.

XX 23-JUN-2000; 2000US-0213849.

XX 16-JUN-2000; 2000US-0228534.

XX 07-SEP-2000; 2000US-0230732.

XX 07-FEB-2001; 2001US-0266862.

XX (SENO-) SENOMYX INC.

XX Zozulya S;

XX WPI: 2001-570867/64.

XX P-PDB: AAU24630.

XX Nucleic acids encoding human olfactory G protein-coupled receptors,

XX useful for screening for compounds involved in olfactory sensation,

XX where the compounds can be used in the food, pharmaceutical and

XX cosmetic industries to customise odours.

XX Claim 1; Page 135; 319pp; English.

XX The invention relates to nucleic acids encoding human olfactory

XX receptors, OR, (a G protein-coupled receptor, GPCR). The OR's

XX specifically recognise molecules, odourants, that elicit specific

XX olfactory sensation. The human olfactory receptors and polynucleotides

XX encoding them are useful for screening a library of chemical compounds

XX for compounds that are involved in olfactory sensation. Modulators of

XX their activity are useful for pharmacological and genetic modulation of

XX olfactory signalling pathways. Therefore, they can be used in the food,

XX pharmaceutical and cosmetic industries to customise odours and

XX the invention.

XX Sequence 924 BP; 219 A; 235 C; 184 G; 286 T; 0 other;

alignment_scores:

Quality: 1574.00

Ratio: 5.127

Percent Similarity: 100.000

Percent Identity: 99.349

alignment_block:

US-09-975-308-9 x AA542323

Align seg 1/1 to: AA542323 from: 1 to: 924


```

1 MetAsnHisSerValValThrGluPheIleIleLeuGlyLeuThrLys 17
1 ATGAATCAGCGCTGTGTAAGTTCATTTATCTGGGCTCACAAAAA 50
17 sProGluLeuGlyIleIlePheLeuPhePheLeuIleValTyrLeu 34
51 GCCTGAAGTCAGGAAATATCTTCCCTTTTTCATGCTATCTATCTG 100
34 aAlaPheLeuGlyAsnMetLeuIleIleAlaIleValIleTyrSer 50
101 TGGCTTTCTCGGCAACATGCTCATCATTCGCAAAATCTATACAC 150
51 ThrLeuHisThrProMetTyrValIlePheLeuThrLeuAlaVal 67
151 ACCTTCATACGCCCATGTAATGTTTCTTCTGACACTGGCTGTGGA 200
67 PileIleCysThrThrSerIleIleProLysMetLeuGlyThrMet 84
201 CATCATCTGCACAAACAGATCATACGAAAGTGGGGACCATGCTAA 250
84 hSerGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPhe 100
251 CATCAGAAATACCATTCATATGACGGTGCATGTGCCAGCTCTTCT 300
101 PheThrTyrSerLeuGlyValGluMetValLeuPheThrThrMet 117
301 TTCACATGCTCTGGAGCTGAGATGTTCTTCCACCAACAGGCTTA 350
117 rAspArgTyrValAlaIleCysPheProLeuHisTyrSerThrIle 134
351 TGACCGCATATGAGCCATTGTTCCCTCTTCATACAGTACGTATGA 400
134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIle 150
401 ACCACCATATGTGTAGCTTGTCTCAGCATGGTCTATGTCAGTGC 450
151 ThrAsnSerTyrValHisThrAlaLeuIleMetArgLeuThrPhe 167
451 ACCAATCTCTGGGTGCACACAGCTTTCATCATAGGTTGACTTCT 500
167 yProAsnThrIleAspHisPhePheCysGluIleProProLeuLe 184
501 GCCAAACACATGACACACTTCTTGTGATGATACCCCATCTCTG 550
184 eSerCysSerProValArgIleAsnGluValMetValTyrValAl 200
551 TGTCTGTAGCCCTTAAAGATCAATGAGTGTGATGTGTGAT 600
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyr 217
601 ATTACCCCTGGCATAGGGACTTATCTTACCTGCATCTCCATGTT 650
217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyLysA 234
651 TATCATTTGCTATCTCCGATCCGACAGTAGAAGCAAGAGAAGG 700
234 laphSerThrCysSerSerHisLeuThrValValThrLeuTyrSe 250
701 CCTTCAACATGCTCATCTCATCTCACAGTGTGACCTTACATCT 750
251 ProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPhe 267
751 CCGTATATCTACACCTATATCGCCCTGCTCCAGCTATACATTTA 800
267 gAspLysValValAlaIleLeuTyrThrLeuValThrProThrLe 284
801 AGACAAAGGTGAGTACACTATCTGACTCTGACTCCACATTAACC 850
284 rMetValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArg 300
851 CGATGTGTACAGCTTCAGAAATAGGAGATGAGGCAATTAAGGAG 900
301 ValPheAlaPheLeuLysHis 307

```

```

|||||
901 GTGTTTCATTTTCGAAACAC 921
seq_name: /SIDSI/gcgcdata/geneseq/geneseq-emb1/NA2002.DAT:AA517179
seq_documentation_block:
ID AA517179 standard; cDNA; 798 BP.
XX
AC AA517179;
XX
DE 14-FEB-2002 (first entry)
XX
Human cDNA encoding a novel G protein-coupled receptor, NOV6.
XX
Human; ss; G protein-coupled receptor; GPCR; NOV6; cardiant;
KW antiarteriosclerosis; antibacterial; virucide; fungicide; protozoacide;
KW antidiabetic; cytotoxic; analgesic; antispasmodic; nootropic;
KW neuroprotective; antiinflammatory; gene therapy; transgenic animal;
KW cardiomyopathy; atherosclerosis; infection; pain; anorexia; bulimia;
KW asthma; neurological disorder; Parkinson's disease; stroke;
KW Alzheimer's disease; multiple sclerosis; lesionial psoriatic skin;
KW ischaemia; cirrhotic hepatitis; acute pancreatitis; diabetes; cancer;
KW angiogenesis; obesity; olfactory disorder; chromosome 11.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..23 /*tag= a
FT CDS 24..782 /*tag= b
FT FT /*product= "NOV6"
FT 3'UTR 783..798 /*tag= c
FT FT
FT PN
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11901.
XX
PR 11-APR-2000; 2000US-195994P.
PR 11-APR-2000; 2000US-196538P.
PR 26-APR-2000; 2000US-199902P.
PR 26-APR-2000; 2000US-199964P.
PR 27-APR-2000; 2000US-199948P.
PR 27-APR-2000; 2000US-199956P.
PR 27-APR-2000; 2000US-200176P.
PR 17-JUL-2000; 2000US-218995P.
PR 25-JUL-2000; 2000US-220644P.
PR 04-JAN-2001; 2001US-259641P.
PR 29-JAN-2001; 2001US-264851P.
PR 14-FEB-2001; 2001US-268567P.
PR 13-MAR-2001; 2001US-0220644.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shenoy S, Gangoli EAG, Rastelli L, Smithson G, Padigaru M;
PI Vernet CAM, Wolenc AR, Casman SJ, Tchernev VT, Szekeres ES;
PI Gorse W, Alsobrook JP, Burgess CE;
XX
DR MPI: 2002-041291/05.
DR P-PSDB: AAU11099.
XX
PT Human G-protein coupled receptors, NOV1-12, useful for diagnosis and
PT treatment of e.g. cardiomyopathy, also in screening for specific
PT modulators -
XX
PS Claim 11; Page 30; 171pp; English.
XX
CC The invention relates to Human G-protein coupled receptor (GPCR)
CC polypeptides, designated NOV1-12, nucleic acids encoding them,

```


CC their activity are useful for pharmacological and genetic modulation of
CC olfactory signalling pathways. Therefore, they can be used in the food,
CC pharmaceutical and cosmetic industries to customise odours and
CC fragrances. The present sequence encodes a human olfactory receptor of
CC the invention.

Sequence 987 BP; 190 A; 309 C; 235 G; 253 T; 0 other;

alignment_scores:

Quality:	861.00	Length:	306
Ratio:	3.337	Gaps:	0
Percent Similarity:	84.314	Percent Identity:	49.673

alignment_block:

US-09-975-308-9 X AAS42284 . .

Align seg 1/1 to: AAS42284 from: 1 to: 987

[illegible][illegible]

seq_documentation_block;

ID AAD12952 standard; CDNA; 1241 BP.

AC AAD12952;

DT 16-OCT-2001 (first entry)

DE Human G-protein coupled receptor-9 (GCRC-9) cDNA.

KM Human; protein coupled receptor-9; GPCR-9; glycolytic; hepatotropic;
 KM virucide; antiinflammatory; anticonvulsant; antileptic; neuroprotective;
 KM neuroptic; cerebroprotective; hypotensive; tranquilliser; vulnerary;
 KM ophthalmic; arterial; cell proliferative disorder; actinic keratosis;
 KM anorectic; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
 KM psoriasis; cancer; neurological disorder; stroke; Alzheimer's disease;
 KM Huntington's disease; Parkinson's disease; cardiovascular disorder;
 KM epilepsy; hypertension; varicose vein; vasculitis; dysphagia; dyspepsia;
 KM anorexia; gastrointestinal disorder; pancreatitis; autoimmune disorder;
 KM Addison's disease; Crohn's disease; acquired immune deficiency syndrome;
 KM AIDS; uveitis; infection; trauma; metabolic disorder; diabetes; obesity;
 KM osteoporosis; transgenic animal; gene therapy; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	CDS	255..1241
----	-----	-----------

FT /product="Human GCRC-9 protein"

PN WO200157085-A2

PD 09-AUG-2001.

PF 01-FEB-2001; 2001WO-US03455.

PR 02-FEB-2000; 2000US-0180093.

XX

XX
XX

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

DR P-PSDB; AAE06759.

Novel isolated human G-protein

cardiovascular, gastrointestinal

XX WPI: 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -

PS Claim 8; Page 289; 1857pp; English.

XX The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.

XX Sequence 924 BP; 170 A; 293 C; 219 G; 242 T; 0 other;

alignment_scores:

Quality: 860.00 Length: 305
Ratio: 3.346 Gaps: 0
Percent Similarity: 84.262 Percent Identity: 49.836

alignment_block:

US-09-975-308-9 x AAH31706 ..

Align seg 1/1 to: AAH31706 from: 1 to: 924

2 AsnHisSerValValThrgluPheIleIleLeuGlyLeuThrLysLysSpr 18
10 AACCAAGACGTTGGTAAACGAGTTCATCGACAGCGCTTTGGAGACCC 59
18 OGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValTyrLeuValA 35
60 AGAATAACCGGGGTCTCTATTCAGCTGTTCCTCTCTCTACTCTGCGG 109
35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAnthr 51
110 CCTCAGAGGATATGCTCTCATCCTTGCGCATTCACGTTCAACCTGGG 159
52 LeuHisThrProMetTyrValPheLeuLeuThrLeuAlaValAlaSpr 68
160 CTCACAGCGCTCATGATGTTTCTTACCACTTGCTGCTACATGACAT 209
68 IleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThr 85
210 TATCTGCACCTCTCTCATCATGCCCAAGCGCTGCGCAGTCTGCTCG 259
85 eArgLysAnthrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeu 101
260 AAGAGAGCTCCATCTCTACGGGGGCTGATGGCCAGCTCTATTTCTCT 309
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyr 118
310 ACGTGGCGGCTCATCGTCAGAGCTGCTGCTCAGCGTATGCGCATGA 359
118 PArgTyrAlaAlaIleCysPheProLeuHisTyrSerThrIleMetAsn 135
360 CCGGACGAGGACCATCTGCCCGCTGCATTAACACACATGATGAGCA 409
135 LsHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
410 AGGTGTTCTGACAGCGGCTGCGCAGCGGCTGCTGCTGCGCGCTC 459
152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168

460 AACACGGCCATCCACACGGGGCTTATGCTGCGCTTGATTTCTGCGCC 509
168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeu 185
510 CAATGCTATATCCATTTCTTCTGAGGAGGCCCTCCCTGCTGCTCTCT 559
185 eCysSerProValArgIleAsnGluValMetValTyrValAlaAspIle 201
560 CCTGCACCTCCACCTACCTACAGCGTGCATGATGTGCTGCGGAGTCT 609
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPhe 218
610 TTTCACGCGCATAGGACTTCTCATATACCATCGCCGCTCATGCTTCAT 659
218 eIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysAla 235
660 CGTCTCCAGCATCTCTGAAGTGAAGACTGCTGGGAGGACGAAACCT 709
235 hSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSerPro 251
710 TCTCCACCTCTCTCCACCTCACCGTGGTGCATGATATACACCGCT 759
252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArg 268
760 GTCTTTCACGCTCATATACAGCCGCTCTGCGCTACAGCGGAGAGAG 809
268 PlysValValAlaAlaLeuTyrThrLeuValThrProThrLeuAsnPro 285
810 CAAGTGGCGCGCTGCTGTACACTGTGCTGAGCTCACCTCAACCCCTC 859
285 eValTyrSerPheGlnAsnArgLysGluMetGlnAlaGlyIleArgLysVal 301
860 TCATCTTACTTTCAGAAACAGAGAGGTCAAGACGCTTCAGAGACTT 909
302 PheAlaPheLeuLys 306
910 TTCCCTTTCTTCAGA 924
seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAS42455
seq_documentation_block:
ID AAS42455 standard; CDNA; 954 BP.
AC AAS42455;
DT 18-DEC-2001 (first entry)
XX Human cDNA encoding olfactory receptor AOLFRL30B.
DE Human; olfactory receptor; G protein-coupled receptor; GPCR; odourant;
KW ss; food additive; cosmetic; fragrance; pharmaceutical additive.
OS Homo sapiens.
XX WO200168805-A2.
XX 20-SEP-2001.
XX 13-MAR-2001; 2001WO-US07771.
XX 13-MAR-2000; 2000US-0188914.
XX 24-MAR-2000; 2000US-0192033.
XX 12-APR-2000; 2000US-0198474.
XX 24-APR-2000; 2000US-0207933.
XX 26-MAY-2000; 2000US-0207702.
XX 23-JUN-2000; 2000US-0213849.
XX 16-AUG-2000; 2000US-0226534.
XX 07-SEP-2000; 2000US-0230732.
XX 07-FEB-2001; 2001US-0268662.
XX (SENO-) SENOMYX INC.
XX PA

PI zozulya S;
 XX WPI: 2001-570867/64.
 DR P-PSDB: AA024762.
 XX
 XX Nucleic acids encoding human olfactory G protein-coupled receptors,
 PT useful for screening for compounds involved in olfactory sensation,
 PT where the compounds can be used in the food, pharmaceutical and
 PT cosmetic industries to customise odours -
 XX
 XX Claim 1: Page 194: 319pp; English.
 PS
 CC The invention relates to nucleic acids encoding human olfactory
 CC receptors, OR, (a G protein-coupled receptor, GPCR). The OR's
 CC specifically recognise molecules, odourants, that elicit specific
 CC olfactory sensation. The human olfactory receptors and polynucleotides
 CC encoding them are useful for screening a library of chemical compounds
 CC for compounds that are involved in olfactory sensation. Modulators of
 CC their activity are useful for pharmacological and genetic modulation of
 CC olfactory signalling pathways. Therefore, they can be used in the food,
 CC pharmaceutical and cosmetic industries to customise odours and
 CC fragrances. The present sequence encodes a human olfactory receptor of
 CC the invention.
 CC
 XX Sequence 954 BP: 224 A; 241 C; 182 G; 307 T; 0 other;
 SQ
 alignment_scores:
 Quality: 731.00 Length: 300
 Ratio: 3.137 Gaps: 0
 Percent Similarity: 77.667 Percent Identity: 45.000
 alignment_block:
 US-09-975-308-9 x AAS42455 ..
 Align seg 1/1 to: AAS42455 from: 1 to: 954
 2 ASNHSSerValValThrGluPheIleIleLeuGlyLeuThrLysLysPr 18
 13 AATCAAAACAGCTCCATCGAATTCATCATCTTGGGGTTCGACCACTGAA 62
 18 OGLeuGlnGlyIleIlePheLeuPheLeuIleValIleVal 35
 63 TGAATTCGAGTATTACTCTTACCACATCTTCTTCTGACCTACATATGCA 112
 35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleTyrSerAsnThr 51
 113 CTTTAGAGGCAATGTTTATCATTTGGTGACCATAGCTGATCCAC 162
 52 LeuHisThrPrometLysValPheLeuLeuThrLeuAlaValAspIle 68
 163 CTACACACACCATGATATTTCTTCTAGAAATCTTGGCCCTTATTTGACAT 212
 68 eIleCyThrThrSerIleIlePheIleIlePheLeuGlyThrMetLeuAsn 85
 213 CTGCTACACTACTACTAATGTCCCCAGATGAGTGCATCTTCTGTACAG 262
 85 eGluAsnThrIleSerTyrAlaGlyCysMetSerGlnLeuPheLeuPhe 101
 263 AAGAAAGAAATCATTTCCATGTGAGGCTGTGACCAAGCTTCTTGCATTC 312
 102 ThrThrSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaTyrAs 118
 313 ATTTCCTTTGCTGGCTCAGAGTGTCTCCCTGCGAGCAATGGCAATGTA 362
 118 PArgTyrAlaIleIleCysPheProLeuHisTyrSerThrIleMetAsn 135
 363 TCGATATATTGCTATCTGTAAGCCGTTAAGGATCATTTATATATGAAACA 412
 135 IHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
 413 AGGCCCTGTGACAGCTGTATGACAGCTCATGCTGACATGTGGTTTCTC 462

152 ASNHSTrpValHisThrAlaLeuIleMetAlaArgLeuThrPheCysGlyPr 168
 463 AACTCAGTGTTCACACCCGTTCTGACCTTCCACCGCCCTCTGTGTAA 512
 168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaAsn 185
 513 CAATCAGATCAATATTTCTTCTGTGACATACCTCCCTTGCATCTGT 562
 185 eCysSerProValArgIleAsnGluValMetValTyrValAlaAspIle 201
 563 CTTTGCTGATACCTTCCCATGAAGCTGCTTCCGTCATGGAGATC 612
 202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPheIle 218
 613 CTGATACAGTGCAGCTTCTTCTGTCATCATACCTTCTTCTTACT 662
 218 eIleValAlaIleLeuArgIleAlaThrValGluGlyLysArgLysAlaP 235
 663 CATCTCCACATCTGAGATCGCTTCTCTGAGGGAGGACCAAAAGCT 712
 235 hSerThrCysSerSerHisLeuThrValValThrLeuTyrTyrSerPro 251
 713 TTTCCACAGTGTGCTCCACCTGCTCATGTATTTCTATATATGACAGT 762
 252 ValIleTyrThrTyrIleArgProAlaSerSerTyrThrPheGluArgAs 268
 763 GCTATCTTCACGATGTAGAGCCCATCTCATCTTACTCTTACAGAAAGA 812
 268 PLYSValAlaAlaLeuLeuTyrThrLeuValThrProThrLeuAsnProM 285
 813 TAGATTGATCTCAGTGTGTATAGTGTGTGCACACCATGCAATTCG 862
 285 eValTyrSerPheGlnAsnArgGluMetGlnAlaGlyIleArgLysVal 301
 863 TAAATTATACGCTAAGCAATAGGACATCAAGAGGCTGTGAAGCCATA 912
 seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NN2001A.DAT.AAF61132
 seq_documentation_block:
 ID AAF61132 standard; cDNA; 3600 BP.
 AC AAF61132;
 DT 17-MAY-2001 (first entry)
 XX Human OLFXY cDNA.
 KW OLFXY: gene therapy; olfactory receptor; human; ds.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT CDS 1179..2144 /*tag="a
 FT /product="OLFXY"
 PN DE19937839-A1.
 XX
 PD 15-FEB-2001.
 XX
 XX 13-AUG-1999; 99DE-1037839.
 XX
 XX 13-AUG-1999; 99DE-1037839.
 XX
 XX (BRUE/) BRUESS M.
 XX (BOEN/) BOENISCH H.
 XX Brueess M, Boenisch H;
 XX
 XX WPI: 2001-184006/19.
 XX P-PSDB: AAB46999.
 DR
 XX New human gene OLFXY, encoding an olfactory receptor, useful for

PT diagnosis, treatment and development of pharmaceuticals -
 XX
 PS Disclosure; Page 3; 4pp; German.
 XX
 CC This invention describes a novel human OLFYX gene (I) and its encoding
 CC protein. The invention also describes (1) transcription factors, RNA
 CC polymerases, pharmaceuticals and chemicals that modulate (I), positively
 CC or negatively; (2) mRNA, and its splice variants or isoforms, transcribed
 CC from (I); (3) cDNA derived from the mRNA of (2) or from intron-less
 CC genes; (4) protein (II) derived from the mRNA, (I) or cDNA; (5)
 CC monoclonal antibodies or antisera directed against the whole of (II) or
 CC or more of its epitopes; (6) expression systems that produce native
 CC or recombinant (II); (7) ligand-binding studies or screening assays using
 CC native or recombinant (II) or cells, or cell membranes, that contain
 CC (II); (8) transgenic and knockout animals that express (II) at altered
 CC levels or not at all; (9) gene therapy methods which use (II), (I) or its
 CC related mRNA or DNA; (10) (anti)sense oligonucleotides derived from (I);
 CC and (11) diagnosis and treatment of diseases in which (II) is directly,
 CC or indirectly, involved. (I) encodes an olfactory receptor. (I), and its
 CC derived proteins, mRNA, cDNA and antisense, or sense, oligonucleotides
 CC are useful for treatment and diagnosis of diseases in which (I) is
 CC directly, or indirectly, involved, and for development and evaluation of
 CC new pharmaceuticals or technologies.
 CC
 XX
 SQ Sequence 3600 BP; 1119 A; 655 C; 588 G; 1238 T; 0 other;

alignment_scores:
 Quality: 728.00 Length: 300
 Ratio: 3.085 Gaps: 0
 Percent Similarity: 78.667 Percent Identity: 45.000

alignment_block:
 US-09-975-308-9 x AAF61132 ..

Align seg 1/1 to: AAF61132 from: 1 to: 3600

2 AsnHISserValValThrcLuphelelleleuGlyLeuThrlLysPr 18
 1191 AATCAACGCTATTAAGTAAATTCATCATCTGGGATTCCTCAACCTAAA 1240
 18 cgluLeuGnglyllelelePhelePheleuValValLeuVala 35
 1241 TGAATTGCAAGTTTACTATTACCATCTTCTTCTGACTTATTCTGTA 1290
 35 laPheleuGlyAsnMetleullelleleAlaLyslleTySerAsnThr 51
 1291 CTTTGGAGGAAATATATTATCTTGACGACGTGACATGATCCACAC 1340
 52 leuHISThrPrometyValPheleuLeuThrlLeuAlaValaLasp11 68
 1341 CTGCAATACACCTATATATATTTCTAGGAACTGGCTTATATGACAT 1390
 68 eileCystrThrSerlleleleProlLysMetLeuGlyThrMetLeuThrs 85
 1391 CTGCTACACACACACATGTCCTCCAGATGATGCGACCTCTCTCA 1440
 85 eRGLAsnThrIleSerTyraLagLysMetSerGlnLeuPheleuPhe 101
 1441 AGAAAAAAGCATTTCTATGTGGGGTGTGTGTCACACTTTTTCGATT 1490
 102 ThrTrpSerleuGlyAlaGluMetValLeuPheThrlThrMetAlaTyrs 118
 1491 GTTTTCTTTGGATGATCAGAGTGTCTCTACAGACGACATGACATATGA 1540
 118 ParGlyValAlaIleCysPhePheLeuHISTySerThrIleMetAsn 135
 1541 TCGTACATTCGATCTGCAATCTTAAGTATTCAGTTATTCGAGCA 1590
 135 lSHISMetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
 1591 AGGTTCTATGCAATCATTAAGCAGCCATGCTGGGCTGCTGTTTCCTT 1640

152 AsnSerTrpValHISThrAlaLeuIleMetArgLeuThrcPheCysGlyPr 168
 1641 AACTCAGTGTGCATACAGTGTGACATTTGCTGCTGCTTCTGTGCA 1690
 168 cAsnThrIleAspHisPhePheCysGluIleProPheLeuAlaLeuS 185
 1691 CAAATCAGATTAATTAATCTTCTGTGACATCCCTTCTGATCTGTG 1740
 185 eRcysSerProValArgIleAsnGluValMetValTyraValaLasp1e 201
 1741 CTGTGGAACACTTCTGCAATGAGTGGCCTGATTCACCTGGGCTC 1790
 202 ThrLeuAlaIleGlyAspPheleleuThrcysIleSerTyrglyPhe11 218
 1791 TTCTATGTTGGACTCTTCTTGTATGTCATTTCTTACATTTGCAT 1840
 218 eileValAlaIleleuArgIleArgThrlValGlyLysArgLysAlap 235
 1841 AATCTCCACCATCTTGAGATCCAGTCCAGAGGAGGAAAGAAAGCCT 1890
 235 hEserThrcysSerSerHisLeuThrlValValThrlLeuTyTySerPro 251
 1891 TTCTCATATGTGCTCCACCTGCGCATTTGCTTCTTTATGGCAGC 1940
 252 ValIleTyThrTyrlleArgProAlaSerSerTyThrPheGluArgas 268
 1941 GCCATCTTACATATGATGACGCGCCCATCTCACTTACTCATTAAGAAAGA 1990
 268 pLysValValAlaAlaLeuTyThrLeuValThrlProThrlLeuAsnProm 285
 1991 TAGCTTGTTTCAAGTGTGTGACAGTGTGTGTACCCCATGCTAAACCTTA 2040
 285 eTyAlTySerPheGlnAsnArgLLeuMetClnAlaGlyIleArgLysVal 301
 2041 TAAATTACACATGAGGAATTAAGACATCAAAAGAGCTGCAAACTATA 2090
 seq_name: /SIDSL/gcgsdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH31995
 seq_documentation_block:
 ID AAH31995 standard; DNA; 930 BP.
 AC AAH31995;
 DT 30-JUL-2001 (first entry)
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 568.
 KW Human; olfactory receptor; OR; primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation; ds.
 OS Homo sapiens.
 PN WO200127158-A2.
 PD 19-APR-2001.
 XX
 XX 06-OCT-2000; 2000WO-US27582.
 XX
 XX 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 XX (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
 DR WPI: 2001-290713/30.
 XX
 XX New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 8; Page 407; 1857pp; English.

XX The present sequence is one of a number of isolated polynucleotides
 CC which encode polypeptides involved in olfactory sensation. The
 CC polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary
 CC scents and the identification of the odour receptors used to detect
 CC these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour
 CC receptors that are involved in detecting such secondary scents.
 CC This enables the construction of a scent representation (also called
 CC a scent fingerprint or scent profile), which may be used to re-create
 CC and edit scents. Libraries of olfactory receptors are useful for
 CC determining the interaction pattern of a composition with the receptors,
 CC and can be used for determining differences in the olfactory faculties
 CC of different individuals.

Sequence 930 BP; 204 A; 259 C; 178 G; 289 T; 0 other;

Alignment_scores:

Quality: 725.00 Length: 300
 Ratio: 3.021 Gaps: 0
 Percent Similarity: 80.000 Percent Identity: 45.000

alignment_block:

US-09-975-308-9 x AAH31995 ..

Align seg 1/1 to: AAH31995 from: 1 to: 930

2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrIlySpr 18
 13 AACAGCTCATGCTGAGCTGATTTATCTCATGGGATCCAGACGCC 62
 18 OGLeuGlnGlyIleIlePheLeuPhePheLeuIleValIlyLeuVala 35
 63 TCAGATGAGATCATCTCTTCGCTCTCCATCATAGTTACTGCTG 112
 35 IAPheLeuGlyAsnMetLeuIleIleIleAlaIlySerAsnThr 51
 113 ATGTAGTGGGAAATATGTTGATGTTATCCGATTACAAGACACAC 162
 52 LeuHisThrProMetIlyValPheLeuLeuThrLeuAlaValIly 68
 163 CTCACACACCCATGATTTCTTCCTGCAACCTCTCTTGTGACCT 212
 68 eIleCysThrSerIleIlePheIlyMetLeuGlyThrMetLeuThr 85
 213 GGGCTACCTCTGAGCATTTGCCCGAGATGCTGCTACTTCTTACA 262
 85 eGluAsnThrIleSerIlyAlaGlyCysMetSerGlnLeuPheLeu 101
 263 ATCAACAAGTATATCTCTTCGACCTGTCACCCAGCTTGTGCTTT 312
 102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrIleMetAla 118
 313 GTAGGTTTGGGATGCTGAGTCTATGCTCGGACGACCTGCTATGG 362
 118 PArgTyrValAlaIleCysPheProLeuHisIlySerThrIleMet 135
 363 TCGTTTGGGCAATTTGTCAGCCCTCCACATATAGCACTTACATG 412
 135 IShiMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 151
 413 AGCAGGTCTGCTGGCTCATGCTGAGGCTTACCTGGCTGCTAGTG 462
 152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCys 168
 463 AGTTTAGTGGCCACACCTGCTACCTTACAGCTGAGTACTGTGCTTC 512
 168 OAsnThrIleAsnHisPhePheCysGluIleProProLeuLeuAla 185
 513 CAATATCATCATCATTTCTTCTGGCAAAATCCACACCTTGGCCCTCT 562

185 eRysSerProValArgIleAsnGluValMetValIlyValAlaAspIle 201
 563 CTGCTCAGACCTACTAGTATGATCTGCTCTGAGCTGTGGC 612
 202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerIlyGlyPhe 218
 613 TTCATGTAAATCAGACACCATCCATCATCTTCATCTTACCTTAT 662
 218 eIleValAlaIleLeuArgIleArgThrValGluGlyIlyAsnGly 235
 663 CTTTTCATATCATCATCAATGCTTACGCTGAGGCGCTTAAAGCTT 712
 235 heSerThrCysSerSerHisLeuThrValIlyThrLeuIlySerPro 251
 713 TCTCACCTCGCGGCTCATCTCTACTGCAATCCCTCTTATGACCA 762
 252 ValIleIlyThrThrIlyArgProAlaSerSerIlyThrPheGluArg 268
 763 GTCATGTTATGTACCTGAGGCCAATCCACACTCTCTGACCAAGA 812
 268 PLYsValValAlaAlaLeuIlyThrLeuValIlyProThrIleAsnPro 285
 813 CAGTGGGCTCTGTTCTTACACGTTTATCATCCCATGTAAATCCCT 862
 285 eValIlySerPheGlnAsnArgGluMetGlnAlaGlyIlyArgIlyVal 301
 863 TGATCTACAGTTTGGCGGAACAAGATGTGAAGCTCTTCAAAAAGCTA 912

seq_name: /SIDSI/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32092

seq_documentation_block:
 ID AAH32092 standard; DNA; 957 BP.
 XX AAH32092:
 DT 30-JUL-2001 (first entry)
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 665.
 XX Human, olfactory receptor; OR: primary scent determination;
 KW secondary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation; ds.
 OS Homo sapiens.
 XX WO200127158-A2.
 XX 19-APR-2001.
 XX 06-OCT-2000; 2000WO-US27582.
 XX 08-OCT-1999; 99US-0158615.
 XX 24-FEB-2000; 2000US-0184809.
 XX (DIGI-) DIGISCENTS.
 XX (YEDA) YEDA RES & DEV CO LTD.
 XX Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;
 DR WPI, 2001-290713/30.
 XX New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX Claim 8; Page 448; 1857pp; English.
 XX The present sequence is one of a number of isolated polynucleotides
 CC which encode polypeptides involved in olfactory sensation. The
 CC polynucleotides can be used in screening for olfactory agonists and
 CC antagonists. The methods allow for the determination of primary
 CC scents and the identification of the odour receptors used to detect
 CC these primary scents. The methods also enable determination of
 CC secondary scents and the identification of combinations of odour

CC receptors that are involved in detecting such secondary scents.
 CC This enables the construction of a scent representation (also called
 CC a scent fingerprint or scent profile), which may be used to re-create
 CC and edit scents. Libraries of olfactory receptors are useful for
 CC determining the interaction pattern of a composition with the receptors,
 CC and can be used for determining differences in the olfactory faculties
 CC of different individuals.
 CC

XX Sequence 957 BP; 242 A; 220 C; 178 G; 317 T; 0 other;

alignment_scores:

Quality: 723.00 Length: 300
 Ratio: 3.064 Gaps: 0
 Percent Similarity: 78.667 Percent Identity: 44.667

alignment_block:

US-09-975-308-9 x AAH32092 ..

Align seg 1/1 to: AAH32092 from: 1 to: 957

```

2 AsnHisSerValValThrGluPheIleIleuGlyLeuThrIysSpr 18
   |||.....|.....|.....|.....|.....|.....|
13 AATCAACAGCTAATCAATCAATCAATCAATCAATCAATCAATCA 62
   |||.....|.....|.....|.....|.....|.....|
18 ogIleuGlnGlyIleIlePheLeuPheLeuIleValIleuValA 35
   |||.....|.....|.....|.....|.....|.....|
63 TGAATTCAGCTTTTACTATTCACACTCTCTTCTGACTTATTCGTA 112
   |||.....|.....|.....|.....|.....|.....|
35 IapheLeuGlyAsnMetLeuIleIleIleAlaIleIleIleIle 51
   |||.....|.....|.....|.....|.....|.....|
113 CTTGGAGAGAAATATATATATATATATATATATATATATATAT 162
   |||.....|.....|.....|.....|.....|.....|
52 LeuHisThrPrometIleValPheLeuLeuThrLeuAlaValAsp 68
   |||.....|.....|.....|.....|.....|.....|
163 CTGCATACACCTATGATATATTTCTAGAGCACTGGCTTTATTCAC 212
   |||.....|.....|.....|.....|.....|.....|
68 eIleCysThrThrSerIleIleProIleMetLeuGlyThrMetLeu 85
   |||.....|.....|.....|.....|.....|.....|
213 CTGCTACACCAACCAATGATGATGATGATGATGATGATGATGAT 262
   |||.....|.....|.....|.....|.....|.....|
85 eArgIleAsnThrIleSerIleGlyCysMetSerIleuPheLeuPhe 101
   |||.....|.....|.....|.....|.....|.....|
263 AGAATAAAGCATTTCTTATGCGGGGTGTGTTCACTTTTTCATTT 312
   |||.....|.....|.....|.....|.....|.....|
102 ThrTrpSerLeuGlyAlaGluMetValLeuPheThrMetAlaTrp 118
   |||.....|.....|.....|.....|.....|.....|
313 GTTTCTTTGTAGATCAGAGTGTCTCTAATGCAATGCAATGCAAT 362
   |||.....|.....|.....|.....|.....|.....|
118 PATGTIValAlaIleCysPheProLeuHisIleIleIleIleMet 135
   |||.....|.....|.....|.....|.....|.....|
363 TCGTTTCATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 412
   |||.....|.....|.....|.....|.....|.....|
135 IHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaThr 151
   |||.....|.....|.....|.....|.....|.....|
413 AGGTCTATGATCAATATAGCAGCTCATGCGCTGCTGCTGCTGCT 462
   |||.....|.....|.....|.....|.....|.....|
152 AsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCys 168
   |||.....|.....|.....|.....|.....|.....|
463 AACTCAGTGTGATCAGTGTGATCAGTGTGATCAGTGTGATCAGT 512
   |||.....|.....|.....|.....|.....|.....|
168 oAsnThrIleAspHisPhePheCysGluIleProIleLeuAlaLeu 185
   |||.....|.....|.....|.....|.....|.....|
513 CAATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 562
   |||.....|.....|.....|.....|.....|.....|
185 eCysSerProValArgIleAsnGluValMetValIleValAlaAsp 201
   |||.....|.....|.....|.....|.....|.....|
563 CTGTGGAAGAACTCTGCAATGATGATGATGATGATGATGATGAT 612
   |||.....|.....|.....|.....|.....|.....|
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerIlePhe 218
   |||.....|.....|.....|.....|.....|.....|
613 TTCATTGTGTGACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 662
   |||.....|.....|.....|.....|.....|.....|

```

```

218 eIleValAlaIleLeuArgIleArgThrValGluGlyAsnArgIle 235
   |||.....|.....|.....|.....|.....|.....|
663 AATCTCCACCATCTTGTAGATCCATCTCTCAGAGGAGAGAGAAAG 712
   |||.....|.....|.....|.....|.....|.....|
235 heSerThrCysSerSerHisLeuThrValIleThrLeuThrIle 251
   |||.....|.....|.....|.....|.....|.....|
713 TTCTACATGTGCTCCACCTGGCATTGTCTTCTTTATGCGAGC 762
   |||.....|.....|.....|.....|.....|.....|
252 ValIleThrThrIleArgProAlaSerSerIleThrPheGluAsn 268
   |||.....|.....|.....|.....|.....|.....|
763 GCCATCTTACATGATGAGGCGCATCTCACTTCACTTCACTTCA 812
   |||.....|.....|.....|.....|.....|.....|
268 PIVeValValAlaIleLeuThrIleValIleProThrIleuAsn 285
   |||.....|.....|.....|.....|.....|.....|
813 TAGGTGCTTTCAGTGTGTACAGTGTGTGTACCCCATGCTAAGCC 862
   |||.....|.....|.....|.....|.....|.....|
285 eValIleSerPheGlnAsnArgIleMetGlnAlaGlyIleArgIle 301
   |||.....|.....|.....|.....|.....|.....|
863 TAATTTCACATGAGAGATAGAGACATCAAGAAAGCTGCAAAACT 912
   |||.....|.....|.....|.....|.....|.....|

```

seqname: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32067

seq_documentation_block:

ID AAH32067 standard; DNA; 927 BP.

XX AAH32067;

XX 30-JUL-2001 (first entry)

XX Human olfactory receptor polynucleotide, SEQ ID NO: 640.

XX Human; olfactory receptor; OR; primary scent determination;

XX secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation; ds.

XX Homo sapiens.

XX WO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

XX (YEDA) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I,

XX WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory

XX sensation for identifying olfactory agonists and antagonists -

XX Claim 8: Page 437; 1857pp; English.

XX The present sequence is one of a number of isolated polynucleotides

XX which encode polypeptides involved in olfactory sensation. The

XX polynucleotides can be used in screening for olfactory agonists and

XX antagonists. The methods allow for the determination of primary

XX scents and the identification of the odour receptors used to detect

XX these primary scents. The methods also enable determination of

XX secondary scents and the identification of combinations of odour

XX receptors that are involved in detecting such secondary scents.

XX This enables the construction of a scent representation (also called

XX a scent fingerprint or scent profile), which may be used to re-create

XX and edit scents. Libraries of olfactory receptors are useful for

XX determining the interaction pattern of a composition with the receptors,

XX and can be used for determining differences in the olfactory faculties

XX of different individuals.

SO Sequence 927 BP; 209 A; 253 C; 176 G; 289 T; 0 other;

alignment_scores:

Quality: 707.00 Length: 303
Ratio: 2.958 Gaps: 2
Percent Similarity: 78.878 Percent Identity: 42.244

alignment_block:
US-09-975-308-9 x AAH32067 ..

Align seg 1/1 to: AAH32067 from: 1 to: 927

```

2 AsnHisSerValValThrGluPheLeuLeuLeuGlyLeuThrIleuValThr 18
13 AATGAGAGTCCCTAATGATGATTCCTTCTGAGGCTTCACAGACACACC 62
18 ocgLeuGlnGlyIleIlePheLeuPhePheLeuIleValIleValIle 35
63 TCGTCGTGAGGGCTGCTCTGTTGATTTGCTTTCTTCTTCTTCTCTGTA 112
35 IAPheLeuGlnMetLeuIleIleIleAlaIleValIleIleValIle 51
113 CCTCTGTGGGAACTTCACCATATCATCATCATCATCATCATCATCAT 162
52 LeuHisThrProMetIleValIlePheLeuLeuThrLeuAlaValAsp 68
163 CTTCATACCCCAATGATCTTTTCTGACGACACCTCTTCTTCTTCTGAT 212
68 eileCysThrThrSerIleIleProIleMetLeuGlyIleMetLeuThr 85
213 CTGCTTACCTAGTACCTGCTCTCTGAGACCTTGTATTAATGACAAAG 262
85 erGIuAsnThrIleSerIleValIleGlyCysMetSerGlnLeuPhe 101
263 CAAGAAGAGACGACTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 309
102 ThrTrpSerLeuGlyAla...GluMetValLeuPheThrMetAla 117
310 TCTCTGACATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
117 rAspArgIleValIleIleCysPheProIleIleIleValIleMetAla 134
360 GATCGGTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 409
134 snHisMetCysValIleLeuLeuSerMetValMetAlaIleAlaVal 150
410 ACCCAGGCTTGTGCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459
151 ThrAsnSerTrpValIleIleThrAlaLeuIleMetArgLeuThrPheCys 167
460 GCTAGTTCCTTATTCATGCACTTCTTCTGCAATTCCTGCTGCTGCTG 509
167 yProAsnThrIleAspHisPheCysGluIleProIleLeuAlaVal 184
510 CAACCATAGCTGAGCATTTTATTTGCGAAGTACACGCTCTCTCTCAAG 559
184 eueerCysSerProValArgIleAsnGluValMetValIleValAlaAsp 200
560 TGGCTTGTGTGACACCATGTCATGATGATGATGATGATGATGATGATG 609
201 IleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTrpGly 217
610 GTTCTGTTTGTGTCATTCACACGACATCATCATCATCATCATCATCAT 659
217 eileIleValAlaIleLeuArgIleArgIleValGlyIleValArgIle 234
660 CATAAATCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 709
234 IAPheSerThrCysSerSerHisLeuThrValIleThrLeuValThrSer 250
710 CCTTACGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 759

```

```

251 ProValIleThrThrIleArgProAlaSerSerThrThrPheGluArg 267
760 ACCATATCTACGTACCTGCAACCTAGTACAGCATATGCCAGGACCA 809
267 gasPlyValAlaIleLeuThrIleValIleThrProIleLeuAsp 284
810 AGGAGATTTATCTCCCTCTCTCTACACACATGAGTCCCATCTTAATC 859
284 rIleValIleThrSerPheGlnAsnArgGluMetGlnIleGlyIleArg 300
860 CTATCATCTTACTTTAAGGACACAGATATGAAGAAGCTCTGAGGAA 909
301 ValPheAla 303
910 CTCTCTCG 918

seq_name: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAS42407
seq_documentation_block:
ID AAS42407 standard; cDNA: 1053 BP.
AC AAS42407;
XX
XX 18-DEC-2001 (first entry)
DE
DE Human cDNA encoding olfactory receptor AOLF213.
XX
XX ss; food additive; cosmetic; fragrance; pharmaceutical additive.
XX
XX Homo sapiens.
XX
XX WO200168805-A2.
XX
XX 20-SEP-2001.
XX
XX 13-MAR-2001; 2001WO-US07771.
XX
XX 13-MAR-2000; 2000US-0188914.
XX
XX 24-MAR-2000; 2000US-0192033.
XX
XX 12-APR-2000; 2000US-0198474.
XX
XX 24-APR-2000; 2000US-0199335.
XX
XX 26-MAY-2000; 2000US-0207702.
XX
XX 23-JUN-2000; 2000US-0213849.
XX
XX 16-AUG-2000; 2000US-0226534.
XX
XX 07-SEP-2000; 2000US-0230732.
XX
XX 07-FEB-2001; 2001US-0266862.
XX
XX (SENO-) SENOMTX INC.
XX
XX Zozulya S;
XX
XX WPI; 2001-570867/64.
XX
XX P-PSDB; AAU24714.
XX
XX Nucleic acids encoding human olfactory G protein-coupled receptors,
XX useful for screening for compounds involved in olfactory sensation,
XX where the compounds can be used in the food, pharmaceutical and
XX cosmetic industries to customise odours.
XX
XX Claim 1; Page 173; 319pp; English.
XX
XX The invention relates to nucleic acids encoding human olfactory
XX receptors, OR, (a G protein-coupled receptor, GPCR). The OR's
XX specifically recognise molecules, odourants, that elicit specific
XX olfactory sensation. The human olfactory receptors and polynucleotides
XX encoding them are useful for screening a library of chemical compounds
XX for compounds that are involved in olfactory sensation. Modulators of
XX their activity are useful for pharmacological and genetic modulation of
XX olfactory signalling pathways. Therefore, they can be used in the food,
XX pharmaceutical and cosmetic industries to customise odours and
XX fragrances. The present sequence encodes a human olfactory receptor of
XX the invention.

```

XX Sequence 1053 BP; 234 A; 278 C; 205 G; 336 T; 0 other;
SQ

alignment_scores:

Quality:	707.00	Length:	303
Rating:	3.059		

Ratio:	2.958	Gaps:	2
Similarity:	78.878	Percent Identity:	42.244

Ratio:	2.750	Gaps:	2
Percent Similarity:	78.878	Percent Identity:	42.244

```
alignment_block:
```

US-09-975-308-9 X AAS42407 ..

Align seg 1/1 to: AAS42407 from: 1 to: 1053

2 AsnHisSerValValThnGluPheLeuLeuGlyLeuThrIleuThrlsYsPr 18
 136 AATGAGAGTCCCTATGATGGATTATCCCTTGAAGGCTTCTCAACAACACC 185
 18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValIYrLeuValA 35
 186 TCGTGTGGAGGCTGTTCTTGTATTGTTCCTTTTCCTTACCCCTGA 235
 35 IapHeLeuGlyAsnMetLeuIleIleIleIleAlaIleIleYrSerAsnThr 51
 236 CCGTTGTGGGAAACCTCAACCAATATCATCATCATATTCGATGCCCT 285
 52 LeuHisThrPrometIYrValValPheLeuLeuThrIleValIleValAspI 68
 286 CTTCATACCCTCATCTACTTTTCTCTCAACAACCTCTCTTACTGGACAT 335
 68 eIleCysThrThrSerIleIleProlsMetLeuGlyThrMetLeuThrs 85
 336 CTGCTTCACTACTACCTCCCTGCTCCAGACCTTACTTACTTACCTCAAGAC 385
 85 eRgIAsnThrIleSerIYrAlaGlyCysMetSerClnLeuPheLeuPhe 101
 386 CAAGAACAACCATCATTCAGGCTGGTGTGGCGCAACTCTATATT... 432
 102 ThrTrpSerLeuGlyAla...GluMetValLeuPheThrThrMetAlaIu 117
 433 TCTGTGGCACTGGGCTCCACATGAATGATCCCTTGGCTGACATGGCCTT 482
 117 rAspArgIYrValAlaIleCysPhePheProIleuHisYrSerThrIleMetA 134
 483 GGATGGGATACATGCTGTGTGCAAAACCCCTCCACTATGATGATATGA 532
 134 snHisHisMetCysValAlaLeuLeuSerMetValMetAlaIleAlaVal 150
 533 ACCCAAGGCTTTGCCAACAGCGTGGCATATCTCCGTGGCTGAGTGGTTTG 582
 151 ThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPheCysG 167
 583 GCTAGTGTCCATATCCATGCAACCTTTACTCTGCAATGACCTCTCTGTGG 632
 167 yProAsnThrIleAspHisPhePheCysGluIleProlProIleuAlaIle 184
 633 CAACCATAGGCTGGAGCCATTTATTGGCAAGTACCAAGCTTTCTCAAGT 682
 184 euserCysSerProValArgIleAsnGluValMetValIYrValAlaAsp 200
 683 TGGCTGTGTGGACACACACTGCAATGAATGGTCTTTTGTGTTTAA 732
 201 IleThrIleuAlaIleGlyAspPheIleLeuThrCysIleSerTrpGlyPh 217
 733 GTTCTGTTTGTGTATTTCACACAGCACTCATCTCATCTCTATGGCTT 782
 217 eIleIleValAlaIleLeuArgIleArgThrValGluGlyIYsArgIYsA 234
 783 CATACACTCAACCTGCGTGAGAGATCAATCATATGAAGCAAGCAATAAAG 832
 234 IapHiserThrCysSerSerHisLeuThrValValThrLeuTrpYrSer 250
 333 CCTTTCAGACCTGTGCTCCACCTTAAAGTGTGTATTCTATCTATGGC 882

```

251 privalIIleYrThryrIIleagProAlaSerSerYrThrPhegIuar 267
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::
883 ACCATATCTACGTGTAACCTGCAACCTAGTGACAGCTATGCCAGNACA 932
      :::|||||:::|||||:::|||||:::|||||:::|||||:::
267 gaEplysValAlAlaAlaleuYrThrIleValThrProThrLeuansP 284
      : ||| ::::: :::|||||:::|||||:::|||||:::|||||
933 AGGGAAGTTATCTCCCTCTCTTACACACCATAGTGACCCCACTTAATC 982
      :::|||||:::|||||:::|||||:::|||||:::|||||:::
284 rOkelValYrSerPhegIlnAsnArgIluMetGlnalaglyIIleArgIys 300
      ||:|||||:::|||||:::|||||:::|||||:::|||||:::
983 CTTTCATCTATACTTAAAGAACAGCATATGAAGAAGCGCTGAGGAAA 1032
      :::|||||:::|||||:::|||||:::|||||:::|||||:::
301 ValPheAla 303
      :::|||||:::|||||:::|||||:::|||||:::|||||:::
1033 CTTCTCTCG 1041

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH32033

seq_documentation_block:

ID	AAH32033	standard; DNA; 927 BP.

AC AAH32033-

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polynucleotide, SEQ ID NO: 606.

KM Human; olfactory receptor; OR; primary scent determination;
KM secondary scent determination; polypeptide library; odour receptor;
KM scent profile; scent fingerprint; scent representation; ds.

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

	XX
.50040T0	'8007 DEJ 42
	XX

PA (YEDA) YEDA RES & DEV CO LTD.

Bellessen J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI; 2001-290713/30.

AA	PT	New polynucleotides
AA	PT	

XX sensation for identifying olfactory agonists and antagonists -

PS CLAIM 8; Page 423; 185/PP; English: XX

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors and can be used for determining differences in the olfactory faculties of different individuals.

Sequence 927 BP; 201 A; 278 C; 166 G; 282 T; 0 other;

Percent Similarity: 79.333 Percent Identity: 42.333

alignment_block:

US-09-975-308-9 x AAH32140 ..

Align seg 1/1 to: AAH32140 from: 1 to: 948

```
2 AsnHisSerValValThrGluPheIleIleLeuGlyLeuThrLysLysPr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 AATTTACACCACTGAGTCTTATTTCTTATGTTGATTCACAGATTATCT 62
18 oGluLeuGlnGlyIleIlePheLeuPhePheLeuIleValIleuVala 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 ACCCTCAGAGTCCACACTGTTCTTGATTCCTTCGTATATACATTA 112
35 IapheLeuGlyAsnMetLeuIleIleIleAlaLysIleIleYSerAsnThr 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 CTATGTCGGAAATATACCTTAATATCTTAATATTAATATTAATCAAGC 162
52 LeuHisThrProMetIleValIleLeuLeuThrLeuAlaValAlaSpril 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 CTTCAAAATCCCATGATGATATATTTCTTATGCAACTATCTTCTTAGACAT 212
68 eIleCysThrThrSerIleIleProLysMetLeuGlyThrMetLeuThr 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 CAGCTGTTCTACAGCAATCACTCTAAATGCTGCAAACTCTTGGCAT 262
85 eRgiuAsnThrIleSerIleValIleGlyCysMetSerGlnLeuPheLeuPhe 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 CCAGGAAAGCATCTCTCTTATGCTGATGCTGCACTAACAATGTTTCTTC 312
102 ThrThrSerLeuGlyAlaGluMetValLeuPheThrThrMetAlaIleTyra 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 GCTTCTTTGCGATGCTGAGTGCCTTATCTCTGCGAGCAATGCTTATGA 362
118 PargIleValAlaIleCysPheProLeuHisIleYSerThrIleMetAsn 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 CCGTATGAGCATCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 412
135 IShIsmetCysValAlaLeuLeuSerMetValMetAlaIleAlaValThr 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 GGAGAGTCTGCTGCTTCATTTGTTGGCATATTTTCAGTGGAAATGA 462
152 AsnSerThrValHisThrAlaLeuIleMetArgLeuThrPheCysGlyPr 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 ACATCTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 512
168 oAsnThrIleAspHisPhePheCysGluIleProProLeuLeuAlaLeu 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 CAATATCGTCAATCATTTTCTGATATCCACCTCTGCTGCTTAT 562
185 eCysSerProValArgIleAsnGluValMetValIleValAlaAspIle 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 CATGTACAGACCTCAGATCAACACACTGCTGCTTGTGCTTGTGACAGC 612
202 ThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerIleGlyPheIle 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 TGCATCCAGACAGCACTTTGTTGTAATATTTATTTCTTACTTTCAT 662
218 eIleValAlaIleLeuArgIleArgThrValGluGlyLysArgLysAlaP 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 CCTCATCATGCTGTGAGCATCAAGTCTCAGTGCAGAGCAAGCAAAACAT 712
235 heserThrCysSerSerHisLeuThrValValThrLeuIleYSerPro 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
713 TCTCCACTTGTGCTCCACCTCATAGCAGTCACTTATTTATGAGAGCG 762
252 ValIleThrThrIleArgProAlaSerSerIleThrPheGluArgAs 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 CTCTGTTTATGTTACTACAGCCACCACTAGTATTTCCCTAGACACTGA 812
268 pIleValValAlaIleLeuIleThrLeuValThrProThrLeuAsnPro 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
813 TAAGTGTGGCAGTCTTTATCTGTGATTTTCCCATGTTTATCCAA 862
285 eValIleYSerPheGlnAsnArgIleMetGlnAlaGlyIleArgLysVal 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
863 TATTTATAGTTTCAGAAACAAGATGTGAAAAATGCTCTCAAAAAGCTA 912
```

qb_est2:BG068751	+	431.00	730.98	1.8e-31	732		BG068751	H3068H10-3	NIA Mouse J
------------------	---	--------	--------	---------	-----	--	----------	------------	-------------

438 AATCAGACF

```

seq_name: gb_hlc:BC016940
seq_documentation_block:
LOCUS      BC016940                2021 bp      mRNA      linear      HTC 09-NOV-2001
DEFINITION Homo sapiens, Similar to olfactory receptor, family 2, subfamily A, member 4, clone IMAGE:4424116, mRNA.
ACCESSION  BC016940
VERSION    BC016940.1  GI:16877381
KEYWORDS   HTC.
ORGANISM   human.
SOURCE     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2021)
            Strausberg, R.
            Direct Submission
            Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK     NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcdbpx11.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 27 Row: 1 Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, similarity but not identity to protein
This clone has the following problem: frame shifted.
FEATURES             source          location/qualifiers
     source          1..2021
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4424116"
                     /tissue_type="Kidney, hypernephroma"
                     /clone_lib="NIH_MGC_89"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT          470 a          546 c          427 g          578 t
ORIGIN
alignment_scores:
Quality: 633.00      length: 314
Ratio: 2.839         gaps: 3
Percent Similarity: 71.019      Percent Identity: 41.401
alignment_block:
US-09-975-308-9 x BC016940 ..
Align seg 1/1 to: BC016940 from: 1 to: 2021
2 ASNHSSerValValThrcGluphelleleuGlyLeuThrLylylsFr 18
|||||.....|.....|.....|.....|.....|.....|.....|
438 AATGAGAAATGGTCCAGAGTTCCTCTACAGGGATCTTCCTCGGCC 487

```


A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.

8 / 2

[illegible]


```

249 TyrsrProValIleThrThrTyrIleAcrProAlaSerSerTyrhph 265
||||| : : : ||||| ||||| : : :
1085 TATGGGGGGTCCACTTCATGTATGTGCTGCCAGTTCCTTCACAGTCC 1134
265 egluArGaSpLySValAlAlAlAlaLeuTyrThrLeuValThrProThrL 282
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1135 TAGCAAGACCAATATCATCTGTGTTCATATCAATGTGCACACGAGTC 1184
||||| : : : : : : : : : : : : : : : : : : : : : : :
1185 TGAACCCCTCATTTACAGCTGAGAAATAGAGAGCTGATGAGCTGTT 1234
299 ArgLySVal 301
|||||
1235 AGAAGAGTA 1243

seq_name: gb_gss:AQ428256

seq_documentation_block:
LOCUS AQ428256 470 bp DNA linear GSS 24-MAR-1999
DEFINITION CITR1-El-2578F11.TF CITR1-El Homo sapiens genomic clone 2578F11,
DNA sequence.
ACCESSION AQ428256
VERSION AQ428256.1 GI:4496022
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 470)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
Unpublished (1997)
Other GSSs: CITR1-El-2578F11.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2578F11"
/clone_lib="CITR1-El"
/sex="male"
/cell_type="sperm"
/note="Vector: pbeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 119 a 122 c 80 g 147 t 2 others
ORIGIN

alignment_scores:
Quality: 531.00 Length: 109
Ratio: 4.963 Gaps: 0
Percent Similarity: 98.165 Percent Identity: 97.248

alignment_block:
US-09-975-308-9 x AQ428256 ..

Align seg 1/1 to: AQ428256 from: 1 to: 470

1 MetaOmniServeValValThrcgUpheIleIleLeuGlyLeuThrLySly 17

```

seq_name: gb_gss:A2909618

seq_documentation_block:

LOCUS A2909618 797 bp DNA linear GSS 05-MAR-2001

DEFINITION RPTCt-24-222618.TJ RPTCt-24 Mus musculus genomic clone RPTCt-24-222618

ACCESSION A2909618

VERSION A2909618.1 GI:13228563

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 797)

AUTHORS Zhao,S., Niernann,W., Malek,J., Shatsman,S., Akiret,B., Levins,M., Tsegaye,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPTCt-24

JOURNAL Unpublished (1999)

COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

FEATURES

source Location/Qualifiers

1..797

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPTCt-24-222618"

/clone_id="RPTCt-24"

/sex="Male"

/cell_type="Spleen/Brain"

/notes=Vector: pTARBAC1; Site.1: BamHI; Site.2: BamHI;
RPTCt-24 mouse BAC library produced by Pieter de Jong.
The library was cloned in the pTARBAC1 cloning vector at the


```

39 smetleuilellelelalellyserlethleuHsthPro 55
140 ACCGTCATCTCTGCGCATTAAGCATAGACTGCGCCACACCC 189
56 MetlyrValPheleuThrleuAlaValAspilleCysThrTh 72
190 ATGACTCTTCTGCTGACACATGCTCTTGTGAC...AACTGCTCTC 236
72 rSerilleleProlysmetleuglyThrMetleuThSerluasnthri 89
237 CACACAGCGCCCAAGATGCTGCGCAATACATCTCAGACGACCAACCA 286
89 leSerlyrAlaGlyCysMetSerGlnleuPheleuThrTrpSerleu 105
287 TCTCTCTCTGCGCATGCTCATGCAATGATATTTATCATGACCTGCT 336
106 GlyAlaGluMetValleuPheThrThrMetAlaTyraParTyValAl 122
337 GACATGACACATTCCTCTGCTGCTGATGATGCGCATGACCTTGTGCG 386
122 alleCysPheProleuHsthlySerThrIleMetAsnHsthMetCys 139
387 CGTGTGCGCGCCCTTACATTCACACAGCAAGATGACCATGCTGTG 436
139 alaIleleuSerMetValMetAlaIleAlaValThrAsnSerTyral 155
437 CCGTCGCTGCTGATGATGATGCTGCTGCGCAACTGAGTCTGCTG 486
156 HsthAlaIleuIleMetArgleuThrPheCysGlyProAsnThrleas 172
487 CACACCTGCTGATGCTGCTGCTGCTGATGCTGCTGCTGCTGCTGCT 536
172 PhePhePheCysGluIleProPheleuAlaIleuSerCysSerProv 189
537 CCACATCTTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
189 alaIleleuSerMetValMetAlaIleAlaValThrAsnSerTyral 205
587 CACACCTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 636
206 GlyAspPheIleuThrCysIleSerTyrglyPheIleleuAlaIle 222
637 ACCCATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 686
222 eleuAlrIleArgThrValGlu.GlyIysArglyAlaPheSerThrCys 238
687 CCGTGAAGGCTCATCA...CAAGGGAAGATGAAGCCTTCTCCACCTGT 733
239 SerSerHsthleuThrValValThrleu 247
734 GGCCTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760

seq_name: gb_gss:AZ913406
seq_documentation_block:
LOCUS AZ913406 794 bp DNA linear GSS 05-MAR-2001
DEFINITION RPI-24-165C12.TV RPI-24 Mus musculus genomic clone RPI-24-165C12
, DNA sequence.
ACCESSION AZ913406
VERSION AZ913406.1 GI:13232351
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 794)
Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akimov,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPI-24
Unpublished (1999)
Contact: Shaying Zhao

```

```

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igf.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.choi.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 165 row: C column: 12
Seq primer: T7
Class: BAC ends.

FEATURES
    source
        1..794
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone_lib="RPI-24-165C12"
            /clone_1lb="RPI-24"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /note="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI;
            RPI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pPARBAC1 cloning vector at the
            BamHI sites using MboI partially digested male C57BL/6J
            DNA."

BASE COUNT      254 a      170 c      221 g      149 t
ORIGIN
1
...
alignment_scores:
    Quality: 511.50      Length: 267
    Ratio: 2.583      Gaps: 3
    Percent Similarity: 74.157      Percent Identity: 38.202

alignment_block:
US-09-975-308-9 x AZ913406/rev ..

Align seg 1/1 to reverse of: AZ913406 from: 1 to: 794
14 leuThrIstysProGluLeuGlnGlyIlelePheleuPheleuI 30
|||||.....:|||||.....:|||||.....:|||||.....:
792 CTGACCCCAAGTCAAGAGTGAAGTGTGCTCTTCTTCTTCTTCT 743
30 evalTyLeuValAlaPheleuGlyAsnMetleuIleleleAlaIysI 47
|||||.....:|||||.....:|||||.....:|||||.....:
742 GGTGTATGTGACAACTGCTGCTGGAACCTTCATCATGTGTCACAGTGA 693
47 leTySerAsnThrleuHsthThr.PrometlyrValPheleuThrle 63
|||||.....:|||||.....:|||||.....:|||||.....:
692 CCGTGTGATGCTGCGCTTCACACACCCCATGATTTCTTGTGTCAGAGATT 643
63 uAlaValAlaPheleuIleCysThrThrSerIleleProlysmetleu 80
|||||.....:|||||.....:|||||.....:|||||.....:
642 GTCTGTGCTGATATGCTGCTTCTTCTTCATCATGCCCCCAAGGTTTGG 593
80 lyThrMetleuThrSerGluAsnThrIleSerTyrAlaGlyCysMetSer 96
|||||.....:|||||.....:|||||.....:|||||.....:
592 TAGACCTTCTCTCAGACAGAAAGCATCTTTCATCATGGTGTCTGACT 543
97 GlnleuPheleuPheThrTrpSerleuGlyAlaGluMetValleuPheThr 113
|||||.....:|||||.....:|||||.....:|||||.....:
542 CAGATGTTCTTCTTCCACCTTATGAGGAGTGTGATGTGTTTCTGCTGTC 493
113 rThrMetAlaTyraParTyValAlaIleCysPheProleuHsthIstys 130
|||||.....:|||||.....:|||||.....:|||||.....:
492 TGTATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 443
130 erThrIleMetAsnHsthMetCysValAlaIleuSerMetValMet 146
|||||.....:|||||.....:|||||.....:|||||.....:
442 TGACCATCATGACGACAGGCGTTCATTTGGTTAATATGAGGCTCTGCTG 393

```

```
147 AAlaIleAlaValThrAnSerTrpValHisThrAlaIleMetArgLe 163
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
392 GGGGGGGGGCTTGGCCATCCATAGCAGATTTCTCTTCTGACACT 343
163 unhrPheCysGlyProAsnThrIleAspHisPhePheCysGluIlePro 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 TCCATTCTGTGGAGCCCAATGTTGATCTTCTACTGTGATGTCGCC 293
180 rleuLeuAlaLeuSerCysSerProValArgIleAsnGluValMetVal 196
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
292 AGGTATCAAACTTGCTGCTCACAGACATTTGTTGAGCTGCTGAG 243
197 TyrValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThrCys 213
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
242 ATTCCAAATATAGCGCTGCTGCTACTGCTGTTGTTGTTGTTGCT 193
213 eSerTyrGlyPheIleIleValAlaIleLeuArgIleArgTrpValGlu 230
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
192 GTCATAC..ACAGCATTTCTTATGATGCTTATGATCTATTCAGAGAG 146
230 lYlYsArgLysAlaPheSerThrCysSerSerHisLeuThrValAlaThr 246
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
145 GCAGAGAGAAAGCCATCTCCACCTGACCTCCACATCATCTGTGTCAC 96
247 LeuTyrTyrSerProValIleTyrThrTyrIleArgProAlaSerSer 263
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
95 CTGCACCTTGTGCTCCTGCATCTATCTATGCGAGCCCTTCACATGCC 48
263 rThrPheGluArgAspLysValAlaIleAlaLeuTyrThrLeuValThr 279
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|
47 ...CTCCCGACGATAAAGCATCTCTGTGTCACCTTCACAGTCATCTCC 3
```

seq_name: gb_uss:AZ593814

seq_documentation_block:

LOCUS AZ593814 580 bp DNA linear GSS 13-DEC-2000
DEFINITION 1M0405K03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0405K03 R. DNA sequence.

ACCESSION AZ593814
VERSION AZ593814.1 GI:11716004

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 580)
AUTHORS Mammaliar, Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
and Wright,D., Weiss,R., Stokes,R., Tinney,A., von Niederhausen,A.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0405 Row: K Column: 03
Seq primer: CACACAGAAACAGCTATGAC
Class: plasmid ends
High quality sequence stop: 580.
Location/Qualifiers
1. 580
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0405K03"

FEATURES
source

/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD22nv, purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD2 (g147321141gblAF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

alignment_scores:

Quality: 504.00 Length: 179
Ratio: 3.338 Gaps: 0
Percent Similarity: 84.358 Percent Identity: 51.955

alignment_block:

us-09-975-308-9 x AZ593814 ..

Align seg 1/1 to: AZ593814 from: 1 to: 580

```
129 TySerThrIleMetAsnHisMetCysValAlaLeuLeuSerMetVal 145
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1 TACAGCTCTGATGATGAGCCCAAGATGTGGGGGCTGCGCCATGAGTGT 50
145 lMetAlaIleAlaValThrAnSerTrpValHisThrAlaIleMetVal 162
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|
51 ATGTCCATCAGTGTGTGATGATGATGATGATGATGATGATGATGATG 100
162 rGluThrPheCysGlyProAsnThrIleAspHisPhePheCysGluIle 178
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
101 GGCTGTCATCTGTGTCGACCAAGATCATCCACATCTCTCTGTGAGTGT 150
179 ProProLeuAlaLeuSerCysSerProValArgIleAsnGluValMet 195
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
151 CCCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 200
195 tValTyrValAlaAspIleThrLeuAlaIleGlyAspPheIleLeuThr 212
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
201 GACTCTTTTGGAGATTTCTTTTGGAGCGTCATTTTGTCTTACTCT 250
212 ySleSerTyrGlyPheIleIleValAlaIleLeuArgIleArgThrVal 228
    ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
251 TCGTATCTATGCTGCTGATCATGTCACAGATCCTGCGCATGCTGTGCT 300
229 GluGlyLysArgLysAlaPheSerThrCysSerSerHisLeuThrVal 245
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
301 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
245 lThrLeuTyrTyrSerProValIleTyrThrTyrIleArgProAlaSer 262
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
351 CTCTGTGTCATCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 400
262 eTyrThrPheGluArgAspLysValAlaIleAlaLeuTyrThrLeuVal 278
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
401 GCTACAGTCACAGAAAGCAAGTACTCTCATGCTTATATCATGATCGTC 450
279 ThrProThrLeuAsnProMetValTyrSerPheGlnAsnArgGluMet 295
```



```

AUTHORS      Harrington,J.T., Sherf,B.,Rundlett,S., Jackson,P.D., Perry,R.,
              Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
              , Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
              , E., Veloso,N., Kilka,A., Hess,J., Cothren,K., Lo,K., Offenbacher
              , J., Danzig,J., and Ducar,M.
TITLE        Creation of genome-wide protein expression libraries using random
              activation of gene expression
JOURNAL      Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE      21227151
COMMENT      Contact: Scott J. Cain
              Athersys, Inc.
              3201 Carnegie Ave, Cleveland, OH 44115, USA
              Tel: 216 431 9900
              Fax: 216 361 9596
              Email: scaine@atersys.com
              High quality sequence stop: 544.
FEATURES     Location/Qualifiers
             source          1..796
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_lib="Athersys RAGE library"
                           /cell_line="HT1080"
                           /note="See 'Creation of Genome-wide Protein Expression
libraries using Random Activation of Gene Expression',
Nature Biotechnology' in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT   115 a       271 c       189 g       220 t       1 others
ORIGIN
alignment_scores:
    Quality: 503.50           Length: 247
    Ratio: 2.595             Gaps: 5
Percent Similarity: 78.543    Percent Identity: 43.320
alignment_block:
US-09-975-308-9 x BG197640 ..
Align seg 1/1 to: BG197640 from: 1 to: 796
6 ValThrGluPheIleIleLeuGlyLeuThrArgLysProGluLeuIngI 22
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
37 ATGTCTGAATTCATCCTCGTCTGCCTCTCTCGCTCCGCCACCTCAACT 86
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
22 yllellepheleupherleuIIeValTyrlzeuValAlaIleleucLya 39
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 GATGGCTCTCTGCTGTCTGCTGATGACTGTTCACTGCTGTGGACA 136
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
39 smetleuIIellellealylsleytSer...AsnThleuHisThr 54
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
137 AC...CTGTTCATCATGGCCACCAGTGTGAGGAGGACGACGCTCACACG 183
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 ProMetTyValPheleuLeuThreualalaValaIlelleIcysth 71
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 CCATGTAACCTCTTCTGTGGCGTCTCAAGTCCGAGATCCHTAAAC 233
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
71 rTrhSerIleIleProLysMetLeuGlyThrMetleuthreserGuant 88
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 CGTGCCATCATCCCGGCGCATGTGGCCACCTGTGTGTCCACCCAGCGCT 283
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
88 hrIleseryrAlaGlyCysMetSerGlnleuPheleuPheThrrpser 104
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 CCAATGCGCTTCTGGCTGTGGCCAGTCAATGTTTC...TTTCCTTCAGC 330
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105 leuGly...AlaGluMetValleupherThrrhMetAlaTyraSPaGTY 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 TTGGGCTTCAACCAACATCCTCTCTGTCACAGCATGAGCTAGACAGCTA 380
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 rValAlaIllecysPheProleuHisTyrserThrIleMetlaInHshISM 137
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 CGTGCCATCTGGACACCCCTGTGGCTAACAGAGATCAAGACCCACAGG 430

```

```

137  etCysValAlaLeuLeuSerMetValMetAlaIleValAlaThrAsnSer 153
      |||:::|||||:::  ||  :::  :::
431  GCTCCGGCTGGTGGTGGCTGCTCTGGGGCTGGTGGCTGGCCATGGG 480
      |||:::|||||:::  |||  :::  :::
154  TrpValHisThrAlaLeuIleMetArgLeuThrPheCysGlyProAsnTh 170
      |||  |||:::  |||:::|||||:::|||||  :::
481  ATGGTGGTGCATCGCCATTTTCCAACTTCTGTGTGGATTCCAGA 530
      |||  |||:::  |||:::|||||:::|||||  :::
170  rIleAspHisPheCysGluIleProPheLeuAlaLeuSerCys. 186
      |||:::|||||:::|||||:::|||||:::|||||  |||:::||||
531  GATCCAGCATTTTATGTCATGTGCCACCTCTGTGTAAGTTGGCGTGT 580
      |||:::|||||:::|||||:::|||||:::|||||  |||:::||||
187  ..SerProValArgIleAsnGluValMetValTyrValAlaAspIleThr 202
      |||  |||  :::  |||  :::  |||  :::  |||
581  GAATTAATGTCACGAGCTGGCCCTGACGCTGGGCTGGTGTGATCATGTG 630
      |||  |||  :::  |||  :::  |||  :::  |||
203  LeuAlaIleGlyAspPheIleLeuThrCysIleSerTyrGlyPheIleI 219
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
631  GCATGCTGGGGCTGTTTCTCTCTCATCTCTCCATCTCTCATGCTTCAT 680
      |||  |||  |||  |||  |||  |||  |||  |||
219  eValAlaIleLeuArgIleArgThrValGluGluYlyArgYlyAsnAsp 236
      :::  |||  |||  |||  :::  |||  |||  |||  |||  |||
681  GCGCCAGCATCTTGAGAGATCCCTTCGTCGGAAGTGGGAACAAGGCTTCT 730
      |||  |||  |||  |||  |||  |||  |||  |||
236  eThrCysSerSerHisLeuThrValValThrLeuTyrTyr 249
      |||  |||  |||  |||  |||  |||  |||  |||
731  CCACCTGTCTCTCACTTAATTGGGTGATGTGTCACATAT 771
      |||  |||  |||  |||  |||  |||  |||  |||

seq_name: gb_gss:AZ607393

seq_documentation_block:
LOCUS       AZ607393                               740 bp      DNA
DEFINITION  IM042N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M042N15 R, DNA sequence.
ACCESSION   AZ607393
VERSION     AZ607393.1  GI:11729583
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 740)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Baccorn,T., Duval,B., Hamll,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contract: Robert B. Weiss
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0429 row: N column: 15
            Seq primer: CACACAGGAACACGCTATGACG
            Class: plasmid ends
            High quality sequence stop: 740.
            Location/Qualifiers
                1..740
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M042N15"
                /clone_1ib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab-host="E. Coli strain XL10-Gold, TI-resistant, F-"
                /note="Vector: PM042N1; Purified genomic DNA from M.

```

BASE COUNT	276 a	134 c	159 g	171 t
ORIGIN				

alignment_block:
US-09-975-308-9 x A2607393/rev .

Align seg 1/1 to reverse of: AZ607393 from: 1 to: 740

[illegible]

seq_name: gb_gss:AZ353949

seq_documentation_block:

VERSION AZ353949.1 GI:10464961

KEYWORDS

SOURCE
ORGANISM

100

1111

REFERENCES

1000

PTTJ.F

1111

JOURNAL

COMMENT

FEATYPES

FEATURES
SOUR

electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 173 a 149 c 135 g 219 t
ORIGIN

alignment_scores:
Quality: 493.50 Length: 220
Ratio: 2.920 Gaps: 2
Percent Similarity: 76.818 Percent Identity: 44.545

alignment_block:
US-09-975-308-9 x AZ353949 ..

Align seg 1/1 to: AZ353949 from: 1 to: 676

```

76  ProlysmetleuglythrMetleuthrsergluanshrileSertyal 92
   |||||||.....:|||||.....:|||||.....:
18  CCGAAATGATTGTTGACCTTTAGTGAAGAAAGATATATCTATGT 67
   |||||||.....:|||||.....:|||||.....:
92  aglyCysmetsergluLeuphethrTrpSerleuglyalaglum 109
   |||||||.....:|||||.....:|||||.....:
68  GGGGGGATGTA.CAACCTTTGGGGTCATTCTTTGGTGCACGTGAGA 116
   |||||||.....:|||||.....:|||||.....:
109 etvalLeuphethrMetalarArgyryValalalecysphe 125
   |||||||.....:|||||.....:|||||.....:
117 TCTTCATTCTTACTGTCTAGTGTAGATAGATATGTCATTGTGATAA 166
   |||||||.....:|||||.....:|||||.....:
126 ProleuHISTYserThrIleMetasnhishismetCysValalaleuLe 142
   |||||||.....:|||||.....:|||||.....:
167 CCTCCCACTAATAGCATATCATGTGACCGGAGATGCAATAAGATGTT 216
   |||||||.....:|||||.....:|||||.....:
142 userMetValmetalaIlealValIthAsnserTrpValIshThrIal 159
   |||||||.....:|||||.....:|||||.....:
217 GCTCGAAGATGATGATCAGTGGCTTCTTACATTTATTCACAGTGGCTC 266
   |||||||.....:|||||.....:|||||.....:
159 euIleMetArgleuthrPhecysglyProasnthrIleasphishphe 175
   |||||||.....:|||||.....:|||||.....:
267 TTGTGTGTCAGTCCCATTTTGTGGACCAATAGATGATCATATTTTC 316
   |||||||.....:|||||.....:|||||.....:
176 CysguIleProleuLeuAlaleuSerCysSerProvalArgIleas 192
   |||||||.....:|||||.....:|||||.....:
317 TGTGATGTACATCTGTACTGAACTTGCCTGCACCGACACTTACATGT 366
   |||||||.....:|||||.....:|||||.....:
192 nclValmetValIyryValAlaAspIlethrLeuAlalegIyasphe 209
   |||||||.....:|||||.....:|||||.....:
367 TGGGTGTTGTGACAGCCACAGTGGCCACATTTGATGGGAAGTTTG 416
   |||||||.....:|||||.....:|||||.....:
209 leleuthrCysIleSerTrglypheIleIleValalaleuArgIle 225
   |||||||.....:|||||.....:|||||.....:
417 TCATCTGTGTCATCTCATACAGATCATCTCATATCT...CTGAGAAAG 463
   |||||||.....:|||||.....:|||||.....:
226 ArgthrValIgluIylysarIyAlaIpheserThrCysSerSerHis 242
   |||||||.....:|||||.....:|||||.....:
464 CAGTCTCCGAAAGGACGCAAGGCTCTCCATCTGAGATGAGATCTACAT 513
   |||||||.....:|||||.....:|||||.....:
242 urThrValIthrLeuTrpTrpSerProvalIleIyryThrIyIlearg 259
   |||||||.....:|||||.....:|||||.....:
514 TGTGTGTGTCATCTTTTGTGGCCCTGATTTATGATATATATCGCG 563
   |||||||.....:|||||.....:|||||.....:
259 roIaSerSerTrpThrPhegluArgAspIylyValalaleuIyry 275
   |||||||.....:|||||.....:|||||.....:
564 CTGACACT.....ACCTTCTGAGACAGAAAGTGTAGTATATTTTAC 607
   |||||||.....:|||||.....:|||||.....:
276 ThrleuValIthrProthrLeuAsnProMetValIyrySerPheglu 292
   |||||||.....:|||||.....:|||||.....:
608 ACTATATATACCTCCATGCTGAATCTCTATATTTACATCTAGAAATGC 657

```

292 gclumetgin 295
|||||.....
658 AGAAGTTAAAG 667

seq_name: gb_gss:BH069789

seq_documentation_block:

LOCUS BH069789 698 bp DNA linear GSS 18-JUL-2001
DEFINITION RPI-24-331A10.TYB RPI-24 Mus musculus genomic clone
RPI-24-331A10, DNA sequence.

ACCESSION BH069789
VERSION BH069789.1 GI:14889386

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintet,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPI-24
Unpublished (1999)
Other GSSs: RPI-24-331A10.TYB

COMMENT
JOURNAL
Contact: Shanying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 331 row: A column: 10
Seq primer: T7
Class: BAC ends.
Location/Qualifiers

FEATURES
source
1..698
/organism="Mus musculus"
/strain="C57BL/6J"
/DB_xref="taxon:10090"
/clone="RPI-24-331A10"
/clone_1b="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; site: 1: BamHI; site: 2: BamHI;
RPI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 230 a 150 c 169 g 149 t
ORIGIN

alignment_scores:
Quality: 483.00 Length: 220
Ratio: 2.825 Gaps: 0
Percent Similarity: 74.348 Percent Identity: 42.174

alignment_block:
US-09-975-308-9 x BH069789/rev ..

Align seg 1/1 to reverse of: BH069789 from: 1 to: 698

```

66 ValAspIleIleCysThrThrSerIleIleleProlysmetleuglyThr 82
   |||||||.....:|||||.....:|||||.....:
698 GTGACATCTGCTACCTCTGTCATCATCTTCACAGATGCTGCGTGTGCT 649
   |||||||.....:|||||.....:|||||.....:
82 ThrThrSergluanshrThrIleSerTyralaglyCysmetsergluLeup 99

```

```
648 GTGGGACATGGCTCAACCAATTTCTCAAGTTCGGCTGCAGTTCATATCT 599
99 heLeuPheThrTrpSerLeuGlyAlaGluMetValLeuPheThrThMet 115
598 TCCCTTTACCTCTTGTGCTTATGCTACTGCTACTGCTTGGCAATCATG 549
116 AlaTyraPheTyrValAlaIleCysPheProLeuHisTyrSerThrI 132
548 GCCACGACCGCTATGTGGCTGTGCTGCCAACCATTTCTTATGTACACAT 499
132 eMetAsnHisHisMetCysValAlaLeuLeuSerMetValMetAlaIle 149
498 CATGACTGAGAAAGCCCGGTAGCTGTGTAACGTGGGACATATGTGGCTG 449
149 laValThrAsnSerTrpValHisThrAlaLeuIleMetArgLeuThrPhe 165
448 GTTTTTCAGTGGCTTATTCGGACTGTCAAGCCTTCACCTTTCATTC 399
166 CysGlyProAsnThrIleAspHisPheCysGluIleProPheLeu 182
398 TGTGGAAACAATGATCAATTCATATCTGTACTCCCTCCCATTTGTT 349
182 uAlaLeuSerCysSerProValArgIleAsnGluValMetValTyrVal 199
348 AAAACTCGTATGTGGGACAGCTACATTCAAGAGTGGTATATGTTT 299
199 laAspIleThrLeuAlaIleGlyAspPheIleLeuThrCysIleSerTyr 215
298 TTGCCATTTTGTTCATGCCCTGTGCATGTGTGATCTCTGTGCTCTAT 249
216 GlyPheIleIleValAlaIleLeuArgIleArgThrValGluGlyLysAr 232
248 CTGTTTATCTTGTGGCATTTATGCANATCCGCTCAGCTGGAGCGGAGC 199
232 gLysAlaPheSerThrCysSerSerHisLeuThrValValThrLeuTyr 249
198 CAAACTTCTCTACCTGCACCTCCACCTCAGCTGTAGCTCTCTCT 149
249 ySerProValIleTyrThrTyrIleArgProAlaSerSerTyrThrPhe 265
148 TTGTACTCTTATCTTCATGTACTTGAGAGATATATACAGATCAAGTCTCG 99
266 GluArgAspLysValValAlaAlaLeuTyrThrLeuValThrProThrLe 282
98 GAGAGGACAGAGTTGTGTGTGTCTATACAGTGGTGAACCCCATTAAT 49
282 uAsnProMetValTyrSerPheGlnAsnArgGluMetGln 295
48 GAATCCACTCATCTATAGCCTGAGAAATAAAGAGTAAA 9
```

THIS PAGE BLANK (USPTO)

From: Li, Ruixiang
Sent: Thursday, June 20, 2002 9:56 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/975,308

Please do a standard search on SEQ ID NOS: 8 and 9 against both the commercial and interference nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 10E18
Mail Box 10C01
306-0282

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

BEST AVAILABLE COPY
BEST AVAILABLE COPY
BEST AVAILABLE COPY

RECEIVED
JUN 20 2002
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 6/21/02
Date Completed: 6/26/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: X
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)